Action Change Theory: A Reinforcement Learning Perspective on Behavior Change

Article *in* Review of General Psychology · March 2015

DOI: 10.1037/gpr0000029

Ivo Vlaev
The University of Warwick

Review of General Psychology
2015, Vol. 19, No. 1, 69–95

1089-2680/15/$12.00 http://dx.doi.org/10.1037/gpr0000029

Action Change Theory: A Reinforcement Learning Perspective on Behavior Change

Ivo Vlaev
University of Warwick

Paul Dolan
London School of Economics

Traditional theories of behavior change rely mostly on influencing higher-order mental processes as a route to altering deliberate responses, whereas more recent theorizing postulates that interventions can also rely on using contextual cues influencing lower-order processes as a route to changing spontaneous responses. We propose an alternative mechanistic account based on reinforcement learning theory, which utilizes different action control systems in the brain. Therefore, this account works at a different level of analysis and description, which promises to lead to the development of a more general and integrative theory of behavior change. *Reward systems* generate specific affective states that influence behavior via 3 action controllers. *Innate actions* are stereotyped evolutionarily determined responses to stimuli. *Habitual actions* develop through stimulus-response learning without explicit outcome representations. *Goal-directed actions* are based on an explicit model of the structure of the environment, which utilizes computations of action-outcome contingencies. We describe how these mechanisms for action control parsimoniously explain behavior change theories and techniques.

Keywords: behavior change, motivation, decision neuroscience, behavioral economics

Many domains of life require achieving specific behavioral goals for individual well-being and social good. Such goals may include encouraging people not to smoke, asking them to improve their diet, making them exercise more, convince them to practice safe sex, requiring them to use seat belts, follow speed limits, protect the environment, and so on. Many of the important questions of public policy in the 21st century relate to how individuals respond to various kinds of information and incentives aimed to prompt such changes in individual behavior in relation to health. Therefore, a better understanding of how best to bring about desired behavior change is vital.

In the health field, for example, over the last 50 years, behavioral medicine and epidemiology have evolved to identify, explain, and address personal risk factors (Davidson et al., 2003; Heller & Page, 2002; Rychetnik, Frommer, Hawe, & Shiell, 2002). For example, there has been a massive accumulation of evidence that supports the premise that sedentary lifestyles are a primary cause of cardiovascular disease, cancer at certain important sites and numerous other morbidities (Blair et al., 1995; Broman, 1995; Pate et al., 1995a). Similar evidence has been accumulated for other risk factors like dieting, smoking, alcohol consumption, sexual hygiene, and medical self-examination. Therefore, substantial health losses are attributable to lifestyle, particularly among the least well-off in society (Barr, 1987; Uitenbroek et al., 1996) and significant gains in health could be achieved by relatively small changes in the choices people make (Department of Health, 2004). Therefore, despite these five decades of research on how to change behavior-related risk factors, policymakers, and health care professionals are still faced with a short supply of generalizable, effective, and sustainable interventions that have been translated into health promotion practice (Glasgow et al., 2004).

Possibly as a result of this situation, until fairly recently policymakers have tended to model population-level behavior change, across most domains of life, simply on the assumption that people will respond to financial incentives (fiscal measures) and that they fail to make optimal choices because of a lack of information (e.g., see Cecchini et al., 2010). However, this approach leaves a substantial proportion of the variance in behavior, beyond the effect of informed intentions (conscious motivation), to be explained (Sheeran, 2002; Webb & Sheeran, 2006). For example, Sheeran (2002) report a meta-analysis of 422 studies, which implied that changing intentions would account for 28% of the variance in behavior change ($r = .53$). Meta-analyses of correlations between intentions and specific health behaviors have found similar effects in studies of condom use (Sheeran, Abraham, & Orbell, 1999) and exercise behavior (Hausenblas, Caron, & Mack, 1997). However, when Webb and Sheeran (2006) based their meta-analysis only on (47) experimental (i.e., causal, not correlational) studies, the

This article was published Online First March 2, 2015.

Ivo Vlaev, Warwick Business School, University of Warwick; Paul Dolan, Department of Social Policy, London School of Economics.

Correspondence concerning this article should be addressed to Ivo Vlaev, Warwick Business School, University of Warwick, Scarman Road, Coventry, CV4 7AL, United Kingdom. E-mail: ivo.vlaev@wbs.ac.uk

estimated intention-behavior correlation dropped to .18 (i.e., accounting for ~3% of the variance).[1]

To improve the effectiveness of interventions, some researchers and policymakers have turned their hopes to a recent explosion of evidence in behavioral sciences (mostly in cognitive and social psychology and behavioral economics), which shows that human behavior is very susceptible to various subtle changes in the environment (Ariely, 2008; Thaler & Sunstein, 2008). Such "contextual" influences on human choices of action are often beyond intentional control, which is probably why such influences have been neglected by policymakers and public health experts who focus mainly on changing conscious or rational intentions as a route to behavior change. Webb and Sheeran (2006, p. 259) provide evidence that standard intervention models based on changing cognitions, such as beliefs and attitudes, can produce behavior change effects comparable with effects resulting from interventions utilizing contextual influences on automatic behaviors. Current models of behavior change have not fully integrated this evidence yet, even though it promises to improve the effectiveness of population-wide interventions. The main contribution of our article is to propose such an elaborated approach, which is embodied in a conceptual framework that unifies various models of behavior change (our search criteria for selecting the literature included secondary sources and integrative reviews). Therefore, to be clear from the start, this article is about analyzing how interventions work, not about how to generate them.

Routes to Behavior Change

Two general paradigms for population-wide behavior change have emerged over the years. The first type is intervention models that aim to change high-order cognitions such as beliefs and attitudes as a route to influencing deliberate responses. In particular, persuasion and education campaigns aim to change attitudes by relying on reflective processing of the provided information (see Norman, Abraham, & Conner, 2000; Shumaker et al., 2008). The second type is more recent intervention models that aim to influence lower-order mental processes; thus, triggering spontaneous responses, by changing the context or environment within which the person acts. The second route relies mostly on contextual changes to bring about automatic behavior change without necessarily changing underlying higher-order cognitions such as beliefs and attitudes. The distinction between behaviors resulting from internally cued, reflective, and intentional changes versus behaviors resulting from externally cued, automatic, and reactive changes is well summarized by Bargh and Chartrand's (1999, p. 463) review of theories of self-regulation:

> Contemporary psychology for the most part has moved away from doctrinaire either-or positions concerning the locus of control of psychological phenomena, to an acknowledgment that they are determined jointly by processes set into motion directly by one's environment and by processes instigated by acts of conscious choice and will. Thus, the mainstream of psychology accepts both the fact of conscious or willed causation of mental and behavioral processes and the fact of automatic or environmentally triggered processes.

Nowadays *dual process theories* can be found in social, personality, cognitive, and clinical psychology (Barrett, Tugade, & Engle, 2004; Chaiken & Trope, 1999; Evans, 2008; Evans & Stanovich, 2013; Gawronski & Bodenhausen, 2006; Moss & Albery, 2009, 2010; Mukherjee, 2010; Sloman, 1996; Slovic et al., 2002). For example, Anderson's (1993) influential ACT-R theory of cognition, distinguishes between declarative semantic knowledge and automatic production rules that map external stimuli to adaptive behavioral responses (Anderson et al., 2004; Meyer & Kieras, 1997). In clinical psychology, the contextual-change route has taken a substantial share of research, because classical behavior therapy and cognitive–behavioral therapy focus on underlying learning processes and environmental contingencies of reinforcement (Clark & Fairburn, 1997; Wolpe, 1990). In summary, the dual-process assumption is that the conscious mind is effortful and limited in capacity but provides systematic and deeper analysis, whereas the automatic mind processes many things simultaneously outside of conscious awareness but is more superficial and heuristic (e.g., walking and eating a sandwich is automatic, whereas having a conversation with somebody is conscious and reflective).

However, some researchers have challenged this classification based on 'hard' distinctions between two systems. For example, assumptions about the role of consciousness as a marker for reflective processes are difficult to sustain in the face of emerging work on unconscious goal pursuit in much the same way as do conscious motivational influences, yet without the person's knowledge or conscious intention (Bos, Dijksterhuis, & van Baaren, 2008; Dijksterhuis et al., 2006; Dijksterhuis, Chartrand, & Aarts, 2007). Therefore, even complex processes as goal activation (Bargh & Barndollar, 1996) and the imitation of social behavior (Dijksterhuis, Bargh, & Miedema, 2000) can efficiently operate without intention and conscious awareness (Bargh, 1996). Animals can also exercise computations of such goal-directed behaviors (Dayan, 2008, 2009). Some scholars also argue that associative processes can be subjected to conscious goal-directed control if necessary, which means that these processes are not completely inaccessible and uncontrollable (Chater, 2009). Other accounts even avoid regarding the two systems as distinct and suggest that reflective or propositional processes are slower and serial because they are realized in cycles of associative processes, and only some of these operations need to be conscious (Carruthers, 2009).

To avoid many of the descriptive "dualisms" (e.g., conscious-unconscious, reflective-automatic, or propositional-associative), which led to confusion in the field, we adopt a neuropsychological approach to behavior change, because it aims toward a mechanistic, as opposed to purely descriptive, basis for understanding regulation of action. We also integrate insights from decision

[1] The weakness of existing approaches is illustrated by a recent behavioral intervention to increase physical activity in an at-risk group in primary care (Kinmonth et al., 2008), which involved 1-year individualized behavior-change program delivered by trained facilitators in participants' homes or by telephone (the program was designed to change behavioral determinants, as defined by the well-known *theory of planned behavior*; Ajzen, 1991). The primary outcome after 1 year was daytime physical activity, but the intervention was no more effective than a basic advice leaflet promoting physical activity (nevertheless, the control conditions were also engaged in physical, physiological, and psychological assessment, which could have had specific influences on behavior similarly to the intervention). Wadden et al. (2011) report similarly unsuccessful attempt to apply behavioral interventions in the form of brief lifestyle counselling to achieve weight-loss in obese individuals, as compared with the usual care provided by GPs (visits to receive educational materials and briefly discussing weight management such as using a calorie book).

neuroscience (neuroeconomics), which describes computational models that correlate with specific brain structures, and thus, provides a more veridical account of action choice. Accordingly, the proposed account works at a different level of analysis, which could lead to a more general and integrative theory of behavior change. Ultimately, our approach aims to improve the theoretical mapping between theoretical constructs and behavior change interventions.

Action Change Theory

We propose the *action change theory* (ACT) that provides a more elaborated account of how the brain controls behavior (see Figure 1). ACT is based on recent mechanistic approaches in cognitive neuroscience, specifically, reinforcement learning theory which is a formal model of how organisms acquire complex behaviors by learning to obtain rewards and to avoid punishments (Dickinson & Balleine, 2002; Glimcher, Camerer, Fehr, & Poldrack, 2009; Rangel, Camerer, & Montague, 2008; Sutton & Barto, 1998). Reinforcement learning is a theory of motivated action, that is, actions resulting from computation of "value," hence why it has also been used to model how the brain "decides" what actions to take. Therefore, we claim, behavior change requires reinforcement learning as much as changing behavior requires motivation. In addition, to better understand how reward (value) systems work in humans, ACT also borrows insights from theories of motivation in psychology (Strack & Deutsch, 2004; West, 2006) and anthropology (Boyer, 2006; Curtis, Danquah, & Aunger, 2009). Figure 2 provides working definitions for each construct in the model.

The truly "novel" aspects of ACT theory is in the implications for analyzing how therapeutic behavioral interventions are best attained. In contrast to behavioral change theories assuming the existence of a unitary system for estimating the value of alternative actions, models of reinforcement learning posit the existence of three competing systems for action control. ACT specifies how different action control systems, which are embodied in contrasting neural networks and functions, can be independently or jointly activated to achieve behavior change. In this article we describe how those neuropsychological systems for self-regulation provide a different level of analysis of the myriad theoretical constructs and behavior change techniques reported in the literature. In this way, ACT could also help determine what techniques are most effective in specific circumstances (e.g., knowing that a particular behavior is driven by a specific action system), which should enable the design of more effective interventions. This is an important potential contribution, because interventions usually start with a comprehensive causal analysis of behavior, which aims to determine the internal regulatory processes that need to be influenced to trigger the required actions (Abraham & Michie, 2008; Michie, van Stralen, & West, 2011; Shumaker, Schron, Ockene, & McBee, 2008).

Mechanisms for Action Control

Actions of organisms are adaptively controlled by specific neural systems (Glimcher at al., 2009; Rangel et al., 2008; Vlaev, Chater, Stewart, & Brown, 2011). In this section, we explain in more detail how different systems in ACT determine behavior change.

Reward systems. Motivation is often defined as the activation of behavior (Elliott & Dweck, 1988), which is usually driven by the reward system via the three action controllers. Therefore, this system becomes an important target for behavior change interventions. The reward system is responsible for generating core affective states, also known as *drives* and *emotions*, that work as either positive or negative rewards in response to either appetitive or aversive stimuli, respectively (Aunger & Curtis, 2008; Berridge, Robinson, & Aldridge, 2009; Rolls, 2005, 2014). Zajonc (1980) suggests that affective reactions are faster and more automatic than cognitive reactions (e.g., sudden noises can cause fear before people figure out the source of the noise). In decision making, Cohen, Pham, and Andrade (2008) argue that judgments are often evoked by feelings, a process known as the *affect heuristic*, which works when subconscious emotional evaluations are used as the basis of decisions (Slovic, Finucane, Peters, & McGregor, 2002) and before cognitive evaluation takes place (Kahneman, 2003).

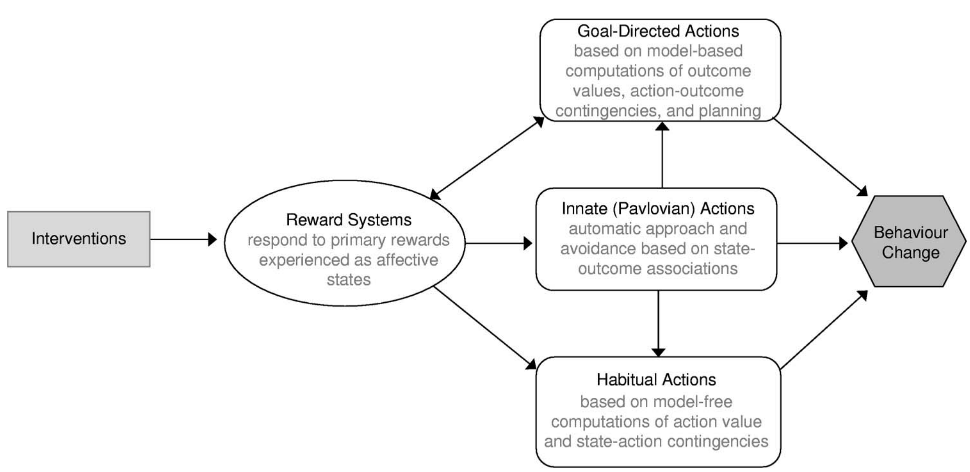

Figure 1. Action Change Theory (ACT) is a reinforcement learning framework for behavior change, which postulates several interactive determinants of action.

Neuropsychological Determinants	Definition	Classification Variables (Computations)	Criteria for Describing and Categorizing Interventions
Reinforcement Learning (RL)	Normative, computational account of action learning and control, which is based on artificial intelligence, and explains how organisms can learn to choose actions that maximize reward and minimize punishments. Reinforcement learning involves four key quantities: states, actions, transitions, and values. The individual has to find a *policy* (choice of action at each state) that optimizes the long-run value of all the values that will be collected.	States	Contexts or stimuli, which can be external or internal.
		Actions	Actions that are available at or given by the states.
		Transitions	Transitions between states that are caused by actions.
		Values	Quantify the immediate value of states in terms of positive or negative reward.
Reward Systems (Values)	The brain systems generating specific affective states which work as *values* for the three types of action (innate, habitual, and goal-directed). The tentative reward types are ordered according being either individualistic (comfort, control, fear, greed, self-enhancement, understanding) or social (attraction, belonging, nurture, status, trusting).	Comfort	Feeling one's body in optimal physical and chemical conditions (proxied by pleasure, pain, effort).
		Control	Feeling contingency between behaviour and outcomes.
		Fear	Feeling risk of injury or death.
		Greed	Feeling in possession of material goods.
		Self-enhancement	Feeling worthy or improvable.
		Understanding	Feeling of shared meaning and prediction.
		Attraction	Feeling attracted to, and attracting, high-value mates.
		Belonging	Feeling of strong, stable relationships and affiliation.
		Nurture	Feeling when caring for offspring.
		Status	Feeling when optimizing social rank.

Figure 2. Definitions for the major classification variables or computations in ACT[a] and criteria for categorizing different intervention effects.

A more elaborate exposition of the possible affective responses would help our understanding of behavior change theories and techniques. Loewenstein (2000, p. 427) suggests that affects such as negative emotions (e.g., anger, fear), drive states (e.g., hunger, thirst, or sexual desire), and feeling states (e.g., pain) are essential in individual daily lives. Fiske (2010, p. 16) provides a comprehensive up-to-date review of human motivations and proposes five core motives, which, we assume, are affective states that function as "primary" (unlearned) rewards (see Rolls, 2014): *belonging* (need for strong, stable relationships, and affiliation), *understanding* (need for shared meaning and prediction), *control* (need for perceived contingency between behavior and outcomes), *self-enhancing* (need for viewing self as basically worthy or improvable), and *trusting* (need for viewing others as basically benign). These core motives, derived from the logic of human adaptation in groups, are used as a theoretical starting point to generate other motives highlighted in the literature (see also Shah & Gardner, 2008). For example, in a systematic meta-analysis of interventions aiming to promote hygiene behavior in 11 developing countries, Curtis, Danquah, and Aunger (2009) subdivides motivated behavior into several categories of drives and emotions that tend to trigger hand-washing behaviors: *affiliation* (seek to conform so as

		Trusting	Viewing others as basically benign.
Goal-directed actions	Actions based on an explicit model or cognitive map of the environment, which is often like a decision tree containing links between states, actions, and outcomes/rewards. Goals depend on the current motivational value of the reward.	Outcome Value	Explicit representation of the expected reward outcomes (costs and benefits).
		Action-Outcome Contingency	The contingency between the action and the outcome.
		Planning	Simulating future possible courses of action and consequent reward outcomes, and searching the decision tree in order to find a good policy.
Habitual actions	Generated as a result of stored past utilities (average rewards) of actions in specific states. Thus the retrospective experience with good and bad outcomes defines present choices without a mental model of the environment. Habits are detached from the current motivational value of the reward.	Action Value	The received reward for taking an action from a given state.
		State-Action Contingency	Repeated experience of an action in a state (also known as Stimulus-Response association).
Innate (Pavlovian) actions	Stereotyped responses expressed as evolutionarily pre-programmed acts, which are emitted on the basis of values associated with a specific state (or state-state associations). Such Pavlovian/classical conditioning differs from instrumental learning (habitual or goal-directed) in the choice of action: automatic action regardless of whether or not it leads to reward, versus learning to emit arbitrary actions based on reward contingency.	Approach	Automatic acts such as locomotion, consumption, fighting, grabbing, biting, scratching, effort, attention.
		Avoidance	Automatic acts such as locomotion, freeze, inhibition, withdrawal, mental avoidance.

[a] For more elaborated definitions see Dolan and Dayan (2013). Also note that the neural models of reinforcement learning in the three action systems suggest that learning the associations between actions and outcomes, and between states and outcomes, depends on neural prediction error signals – indicating the difference between predicted and actual outcomes (see Dayan & Abbott, 2001).

Figure 2. (*continued*).

to reap the benefits of social living), *attraction* (be attracted to, and want to attract, high-value mates), *comfort* (place one's body in optimal physical and chemical conditions), *disgust* (avoid objects and situations carrying disease risk), *fear* (avoid objects and situations carrying risk of injury or death), *nurture* (want to care for offspring), and *status* (seek to optimize social rank). Note that even though the social and physical environments were quite varied across the 11 countries, the specific motivations represent a common universal set (Judah et al. (2009) utilized this framework to develop intervention messages aimed at increasing hand washing in a developed western society). In our discussion of behavior change theories and techniques, we use similar classification of evolved primary rewards that are triggered to influence a specific action controller.

Goal-directed actions. Goal-directed actions require the most complex information processing, because they are based on an explicit model of the structure of the environment. Goal-directed actions require three core computational processes: valuation (costs and benefits) of outcomes (*outcome value*), probabilistic estimation of the contingency between the action and the outcome (*action-outcome contingency*), and *planning* which incorporates those calculations and engages in modeling and searching through decision trees containing sequences of state-action-outcomes to calculate the optimal sequence of actions (i.e., involves representing explicit models of the world or the organism). Therefore, goal-directed action control learns the transition structure of the environment separately from the outcome values (the latter makes goal-directed actions sensitive to the current motivational state of the organism). Persuasive information tends to trigger goal-directed actions as such information is often about state-action-outcome sequences. Such representations can be flexibly generated and changed, but to do so, and to suppress unwanted impulses, the system needs attentional and computational resources (Mann & Ward, 2007).

Habitual actions. Habitual actions are stimulus–response associations learned through repeated practice and rewards in a stable environment. Control over decisions often transfers from goal-directed mechanisms to the habit systems that control *motor habits* and *mental habits*. This distinction is supported in the literature. According to Bargh (1996, p. 28) "any skill, be it perceptual, motor, or cognitive, requires less and less conscious attention the more frequently and consistently it is engaged." Habitual actions are mediated by instrumental learning (sometimes referred to as operant conditioning), whereby an individual learns to associate a particular action (e.g., movement or cognitive strategy) with its value in a given state, without an explicit representation of the specific outcome or goal (that is a privilege of the goal-directed actions). Thus, the two essential computations are estimation of *action value* and *state-action contingency*, respectively. Consequently, motor or mental actions that lead to reward are executed more frequently in the specific state, whereas those that lead to aversive events are executed less often. Even though habitual actions require intentionality and goals to begin, longer-practiced (strong) habits are usually difficult to change even when they are in opposition to intentions, which supports the dissociation between habit and planning systems (Neal, Wood, & Quinn, 2006; Ouellette & Wood, 1998).

Motor Habits are instrumental responses based on adaptive state-action contingencies or associations (also known as operant conditioning), thus, avoiding the need to compute the expected outcomes. Motor habits are usually defined as "behavioral dispositions to repeat well-practiced actions given recurring circumstances" (Wood, Tam, & Witt, 2005), which develop through repetition (e.g., smoking when reading the news) in the presence of consistent states/stimuli (e.g., coffee or home; Neal, Wood, & Quinn, 2006) and rewards (Rangel, Camerer, & Montague, 2008). This process leads to habits being automatically cued by environmental cues and easier to perform over time. Such habits are usually the consequence of past goal pursuit (e.g., relaxation or socializing), but once acquired, habits are cued and performed without mediation of a goal (Wood & Neal, 2007). Wood et al. (2002) estimate that substantial proportion of our everyday activities are performed habitually (e.g., eating, exercising, drinking, driving, and hygiene practices). Semantic concepts can also be automatically connected to habitual action sequences (e.g., Bargh, Chen, & Burrows, 1996; Dijksterhuis & Bargh, 2001). Such state-action (sensory-motor or conceptual-motor) associations have important implications for cueing "good" habits in behavior change interventions, as we discuss later.

Mental Habits are automatic processes that are relatively well-researched in cognitive and social psychology. Automaticity of such cognitive procedures is achieved through frequent execution in response to cues and utilizing connections in long-term memory (Anderson, 1993; Hummel & Holyoak, 2003). Verplanken et al. (2007) investigated negative self-thinking as a mental habit—with a key distinction between *mental content* (negative self-thoughts) and *mental process* (negative self-thinking habit). The latter was assessed with a metacognitive instrument (Habit Index of Negative Thinking) measuring whether such habitual thoughts occur often, unintended, initiated without awareness, difficult to control, and are self-descriptive (e.g., habitual negative body-image thinking is a psychological risk factor in adolescents, Verplanken & Velsvik, 2008). We extend the mental habit concept to include well-documented *heuristics*, such as certain kinds of judgments that either depart from rationality or represent adaptive cognitive strategies that exploit the statistical structures in the environment (Gigerenzer & Selten, 2001; Gigerenzer, Todd, & the ABC Research Group, 1999; Simon, 1992). For example, *lexicographic heuristics* such as *satisficing* (Simon, 1956), *elimination by aspects* (Tversky, 1972) and *one reason decision making* (Gigerenzer & Goldstein, 1996), describe the tendency to make decisions by sequentially using only one choice attribute at a time (e.g., price or familiarity) starting with the most important or salient attribute, and continuing until all unsatisfactory alternatives are eliminated. Computational *reinforcement learning* models predict how such mental habits are acquired and deployed (see Erev & Barron, 2005).

Innate (Pavlovian) actions. Innate actions are evolutionarily appropriate responses to specific predetermined stimuli, although associative learning allows organisms to deploy them in response to other stimuli—such behaviors are also known as unconditioned and conditioned Pavlovian responses, respectively (Mackintosh, 1983). Thus, Pavlovian learning of *state-state* and *state-outcome* contingencies allows organisms to learn the predictive value of a state/cue, which reflects the sum of rewards and punishments expected to occur from it (see Seymour & Dolan, 2008). Such value expectancies can instigate two fundamental types of evolved reactions: *approach* (designed to decrease the distance between the organism and a feature of the environment using responses such as locomotion, grabbing, consumption, fighting, and mental approach/focusing) and *avoidance* (responses aim to increase the distance by moving away, flight, freeze, and mental avoidance). Note that even though some of those specific actions can be used in goal-directed behaviors (e.g., animal defending a held resources may use cost-benefit planning), the signature of innate actions is their automaticity regardless of whether or not they lead to immediate reward in the given situation.

The dissociation, or direct competition, between innate actions and the instrumental (habit and goal-directed) actions is illustrated by animal experiments that set them in opposition. Apparently, animals cannot learn to withhold the innate response that leads to maladaptive or self-destructive behavior when punished for emitting an innate response to that punishment. For example, in a procedure known as *negative automaintenance*, when denying food to pigeons when they peck a key that lights up before it delivers food, birds still often peck the key thereby getting less food despite the instrumental contingency between withholding pecking and food (Williams & Williams, 1969). Likewise, chicks cannot learn to retreat from a food cart that moves in the same direction as them but at twice the speed, and they keep approaching the cart even though this action is never rewarded (Hershberger, 1986). In the same way, squirrel monkeys increase pulling on a restraining leash that delivers painful electric shocks when they do so, instead of doing the optimal response to stay still (Morse, Mead, & Kelleher, 1967); and if Siamese fighting fish are instrumentally punished for their innate aggressive fighting display, the display continues regardless (Melvin & Anson, 1969). In summary, these examples demonstrate that innate actions are formally independent from instrumental actions because the former are never reinforced.

This innate predetermination offers appropriate default actions in specific environments, which offer the advantage not to have to learn what actions to take, particularly in the case of threats (Dayan, 2012, p. 43). However, innate actions can also underpin a surprisingly wide range of human behaviors that have maladaptive consequences. Innate actions can also lead to overeating, addiction, obsessive–compulsive behaviors, and opting for immediate smaller rewards at the expense of delayed larger rewards (see Dayan, Niv, Seymour, & Daw, 2006). This follows from the character of the innate actions, which are myopic and directed toward "primary" (unlearned) rewards and their predictors (Rolls, 2014).

Innate actions can also influence habits and goal-directed actions (see Figure 1), which involves a broad range of phenomena known as *Pavlovian-Instrumental Transfer* (PIT) (see Niv, Joel, & Dayan, 2006; Talmi, Seymour, Dayan, & Dolan, 2008). PIT-affected actions are either goal-directed (*outcome-specific PIT* mediated by amygdala's central nucleus) or habitual (*general PIT* mediated by amygdala's central nucleus). General PIT occurs when stimuli associated with an appetitive/aversive outcomes modulate the vigor of a habitual response reinforced by a different reward, as if the predicted outcome is added to the consequences of the action enhancing/suppressing the desire to perform it (see Cartoni et al., 2013). For example, cues predicting a fun time may increase smoking, while stimuli predicting aversive outcomes can suppress appetitive responding and lead to withdrawal (also known as "conditioned suppression;" Estes & Skinner, 1941). Outcome-specific PIT occurs when the affected actions are goal-directed, which is observed when a Pavlovian stimulus associated with an appetitive/aversive outcome motivates/inhibits an instrumental action leading to the same outcome. For example, environmental drug-associated cues provoke drug-seeking behavior and relapse in recovering addicts (Everitt, Dickinson, & Robbins, 2001), whereas states predictive of aversive outcomes can suppress goal-directed information search (Dayan & Seymour, 2009; Gray, Braver, & Raichle, 2002; Ochsner & Gross, 2005), which explains why individuals tend to contemplate future paths with more positive outcomes and mentally block out paths leading to aversive outcomes (e.g., ignoring reasons to participate in cancer screening, Moser, Patnick, & Beral, 2009). Although outcome-specific PIT is not sensitive to devaluation, as goal-directed actions must be, if the Pavlovian stimuli are removed then the actions often are sensitive to devaluation (i.e., goal-directed), at least when the Pavlovian contingencies are specific (e.g., S1-O1, S2-O2). Balleine and O'Doherty (2010) suggest a model that implies the S-O association selects the R via O-R association, which then also retrieves the goal-directed R-O association (this process may involve a distributed representation of the outcome as both a goal O^G in R-O associations and as a stimulus O^S in S-O and O-R associations).

Neural Implementation

The systems presented in Figure 1 are also implemented in relatively segregated functional neural networks.[2] The idea that the brain contains separate decision systems is ubiquitous in psychology and neuroscience (see Balleine, 2005; Gottlieb & Balan, 2010). What uniquely differentiates the three action systems is their engagement of the rich network of reward systems—the corticostriatolimbic circuits—in the brain (see Seymour, Singer, & Dolan, 2007); and as a result, each system assigns a value to each available action, and thus, competes with the actions favored by the other systems (such competition is implicated in self-control issues such as dieting or drug addiction; Daw, Niv, & Dayan, 2005).

Reward Systems are implemented in a wide-spread network of brain regions involved in processing motivational relevance. Brain structures commonly linked to affect are the amygdala, the cingulate cortex, and the insular cortex (LeDoux, 2000): the amygdala has been implicated in processing motivationally relevant stimuli, valence computation and conditioning (Phelps, 2006; Seymour & Dolan, 2008); anterior cingulated cortex is involved in arousal and assessing the salience of motivational information (Allman et al., 2001); and insular cortex represents somatic information, particularly as it relates to arousal and feelings (Critchley et al., 2004; Damasio, 2000), disgust (Phillips et al., 1997), empathy (Singer et al., 2004), pain and visceral sensations (Critchley, 2005), and interoception that may be involved in decision-making by representing valenced subjective states (Damasio, 1994).

Differentiation between the components of the reward system and the three action systems (as in Figure 1) resonates with recent neurobiological evidence dissecting three dissociable psychological components of reward: "liking" (hedonic impact), "wanting" (incentive salience that provokes approach toward and consumption of rewards), and "learning" (predictive associations between stimuli and reward; see Berridge, Robinson, & Aldridge, 2009). In particular, hedonic hotspots for opioid enhancement of sensory pleasure, that is, "liking," are located in the nucleus accumbens' medial shell (part of ventral striatum) and the target for its outputs—the posterior ventral pallidum. Thus, the dissociation of liking reveals a separate reward system responsible for generating specific affective states. The reciprocal influence between the reward system and the goal-directed system is also established in the literature. Goal-directed actions use affective signals for the hedonic valuation of future states (Ochsner & Gross, 2005), whereas dorsolateral prefrontal cortex is specifically involved in cognitive control over emotions (Delgado, Gillis, & Phelps, 2008; Miller & Cohen, 2001) such as when delaying gratification (McClure et al., 2004, 2007).

Goal-directed actions are learned and implemented in specific regions of frontal cortex (medial prefrontal/medial orbital) and of basal ganglia (dorsal striatum—anterior caudate in humans and

[2] The neural circuitry of all action controllers is more complex than is presented here (see the provided references for more detailed explanations), which is not the focus of the article. Note also that although the ideas about multiple learning and memory systems in the brain were developed and proposed earlier (e.g., Hirsh, 1974; O'Keefe & Nadel, 1978; recent review by White, Packard, & McDonald, 2013), here we focus on the more recent publications directly related to reinforcement learning theory (as a theory of motivation and action).

dorsomedial striatum in animals), but may also subsume mechanisms localized in hippocampus and lateral prefrontal cortex, which mediate declarative expectations of future outcomes and conscious planning (Balleine & O'Doherty, 2010; Berridge & Kringelbach, 2008; Glimcher et al., 2009; Rolls, 2014).

Habits are implemented in specific subcortical, basal ganglia structures—posterior lateral putamen in humans and dorsolateral striatum in animals, and also include dopamine neurons into this area (arriving from substantia nigra and the ventral tegmental area) that are important for learning the value of habitual actions. Stimulus-response representations might also be encoded in cortico-thalamic loops. Zink et al. (2003, 2004) show that the dorsal and ventral striatum's involvement in habitual action selection and reward processing depends on the salience associated with reward and related environmental cues.

There is also evidence about the neural interactions between habits and goal-directed control (also referred to as *model-free* and *model-based* actions, respectively). As described, habits require extensive experience including schedules of reinforcement involving single actions and single outcomes, which implies that behavior must be initially goal-directed and gradually becomes habitual over the course of experience. This view is supported by evidence for the transfer from dorsomedial (caudate) to dorsolateral (putamen) striatum over the course of training (see Dolan & Dayan, 2013). The vulnerability of goal-directed control to "intrusions of habits" (when goal-irrelevant well-trained actions are elicited by contextual cues) is predicted by the gray matter density in the putamen and the strength of white matter connections between premotor cortex and posterior putamen; whereas expressing flexible goal-directed control, such as when responding to changes in the reward value of outcomes, is predicted by the tract strength between caudate and ventromedial prefrontal cortex (de Wit et al., 2012; Dolan & Dayan, 2013).

Innate actions are also implemented in the subcortical structures. Approach and avoidance responses are mediated partially by the action of dopamine and serotonin, respectively (Boureau & Dayan, 2011). Nonspecific preparatory responses are also controlled by the amygdala through its connections to the brain stem nuclei and the core of the nucleus accumbens, whereas more specific responses are controlled through amygdala's connections to the hypothalamus and the periaqueductal gray (innate responses to negative stimuli have specific and spatial organizations along an axis of the dorsal periaqueductal gray).

Summary of the Remaining Sections

In the remaining three sections of this article, we discuss how the elements (systems and processes) in ACT offer a novel, mechanistic interpretation of the major behavior change theories and techniques in the literature. The next section examines the conceptual relationship between ACT and the major theoretical approaches to behavioral change. The subsequent section discusses how specific behavior change techniques selectively influence the three action systems. The final section concludes and outlines open issues for future research.

Here it is crucial to stress two major caveats. First, any one of the interventions discussed later in this article involves multiple processes. Second, even if a single process could be identified for a particular intervention, it is manifest at multiple levels of analysis, including the neuropsychological learning mechanisms we identify, as well as the affective, cognitive, and behavioral manifestations of and contributors to these mechanisms. For example, even though this article focuses exclusively on highlighting the contribution of a single level of analysis, the learning mechanisms that we highlight are potentially compatible with social psychological constructs. Social psychological models are built largely around constructs involving the "self-concept," "social norms," "subjective interpretation," or "construal of experiences." Some of those constructs, such as "interpretation" for example, are not addressed in the manuscript. The discussion of reward systems, for example, covers only primary (unlearned) rewards while excluding some of the literature on motivational theories within the social/personality tradition, such as *self-determination theory* (Deci & Ryan, 2002). However, we know that the ways people construe rewards makes a difference in how willing they are to engage in a particular behavior. Another example is Wood and Neal's (2007, 2009) discussion of the multiple processes involved in habit formation, including individuals' construal of experienced contingency. In contrast, our focus here is on highlighting how the neuropsychological level of analysis might improve our understanding of behavior change in addition to what we already know from other levels of analysis (having said that, we do think that future research should attempt to elaborate on how these sorts of psychological contributions are linked to the neuropsychological mechanisms included in ACT).

A Reinforcement Learning Perspective on Theories of Behavior Change

Psychological theories of behavior in social and health psychology define behavioral determinants, such as beliefs and goals, which can be targeted by interventions producing changes in those mental representations (Sloman, 1996) and ultimately behavior. Three main categories of theories that have been used to inform the design of behavior change interventions: *attitude theories*, *goal theories*, and hybrid *stage theories* (see Webb & Sheeran, 2006). Here we aim to demonstrate that even though these numerous and complex models may look different from each other in some fundamental way, they ultimately assume the underlying processes proposed in ACT. This offers a novel conceptualization of how various constructs from diverse models and research domains link together, but it also highlights what empirically established constructs are missing in some models.

Table 1 summarizes those theories and outlines, according to ACT, the neuropsychological mechanisms underpinning the behavioral determinants postulated by each theory—in terms of similar underlying computations and processes (the terms in the table are easily accessible to readers for reference). For example, planning, a key computation in goal-theories, involves constructing and searching a causal model of the task, such as a decision tree, to work out the value of each action and find the optimal sequence of actions. Thus, the focus here is on "what" is computed (see Dayan & Abbott, 2001, for plausible algorithms specifying "how" those values are computed), because such deeper understanding of the underlying mechanisms that drive behavior should also provide insights into the type of interventions that are likely to engage distinct action systems with potential beneficial effects.

Table 1
ACT Provides the Common Behavioral Determinants for Prominent Theories of Behavior

Behavioral domain	Theory name	Definition	Behavioral determinants in ACT	
			Action	Reward
Attitude theories	Theory of reasoned action (TRA; Fishbein & Ajzen, 1975)	Attitudes (beliefs about likelihood of behavioral outcomes and evaluation of those outcomes) and beliefs about how others value the action determine behavioral intentions.	Goal-directed	Comfort Status
	Theory of planned behavior (Ajzen, 1991)	The same as the TRA plus perceived behavioral control (i.e., action probability).	Goal-directed	Comfort Control Status
	Model of interpersonal behavior (Triandis, 1977)	Behavior is caused by attitudes (beliefs about outcomes and evaluation of outcomes), social factors (norms, roles, self-concept), emotions, motor habits, and facilitating conditions.	Goal-directed Habit	Comfort Control Self-worth Belonging Status
	Health belief model (Rosenstock, 1974)	Preventative health behavior is determined by beliefs about benefits, barriers, vulnerability, and severity.	Goal-directed	Comfort Fear
	Protection motivation theory (Rogers, 1983)	Appraised severity, expectancy of exposure, and belief in efficacy of coping response, jointly determine protection motivation that changes intentions to respond.	Goal-directed	Comfort Fear
	Social support theory (Heaney & Israel, 1997)	Social contacts influence health behavior by providing four types of social support: emotional, instrumental, informational, and appraisal.	Goal-directed	Control Understanding Belonging Status
	Elaboration-likelihood model (Petty & Cacioppo, 1986)	Persuasive messages are processed via two routes: central (involves effortful deliberation depending on motivation and capacity) and peripheral (emotional, less conscious, and relies on contextual cues).	Goal-directed Habit	Fear Attraction
	Heuristic-systematic model (Chaiken, Liberman, & Eagly, 1989)	Persuasive information is processed either in a high-involvement and high-effort systematic (analytic) way, or through shortcuts/heuristics that are unconsciously activated and applied. Information processing is biased by motivation for accuracy, social impression, and/or identity defense.	Goal-directed Habit	Self-worth Understanding Belonging Status
	MODE (motivation and opportunity as determinants) model (Fazio, 1990)	Attitudes directly (not via intentions) guide behavior by a combination of deliberative processes (as in TRA) and spontaneous processes (under pressure accessible attitudes bias definition of events and cause behavior as a heuristic).	Goal-directed Habit	Comfort Understanding
	Prototype-willingness model (Gibbons, Gerrard, Blanton, & Russell, 1998)	Posits two routes to behavior: rational/intentional (based on attitudes and social approval) and an automatic (based on status images, and past behavior).	Goal-directed Habit	Status

(*table continues*)

Table 1 (*continued*)

Behavioral domain	Theory name	Definition	Behavioral determinants in ACT: Action	Behavioral determinants in ACT: Reward
	Reflective-impulsive model (Strack & Deutsch, 2004)	Dual-process theory of social behavior, which integrates cognitive (attitudinal, reasoning), motivational (approach, avoidance), and behavioral (action schemata) mechanisms.	Goal-directed Habit Innate	Comfort Fear
Goal theories (self-regulation theories)	Control theory (Carver & Scheier, 1982)	Self-regulation is an ongoing process of comparing one's performance with a goal (desired state or outcome) and adjusting behavior as a result.	Goal-directed	Comfort
	Social-cognitive theory (Bandura, 1991)	Self-regulation has two elements: self-monitoring provides the contextual information (or reference value) whereas self-judgement sets the target level (as input value, goal setting, or social standard/imitation).	Goal-directed	Control Self-worth Belonging
	Theory of goal setting (Locke & Latham, 1990)	Behavior changes a result of carrying out specific behavioral tasks as a way to achieve a more general goal.	Goal-directed	Control
Stage theories (integrating aspects of attitude and goal theories)	Model of action phases (Heckhausen & Gollwitzer, 1987)	Emphasizes the temporal aspect of goal pursuit. Thus, the predecisional phase is the first step in behavioral change, in which the individual contemplates the feasibility and desirability of various goals and selects the one(s) to pursue.	Goal-directed	Comfort Control
	Transtheoretical model (Prochaska & DiClemente, 1984)	Goals change adaptively according to the person's motivational readiness for change. People process through five stages of change in their pursuit and attainment of short and long term goals: precontemplation, contemplation, preparation, action, and maintenance.	Goal-directed	Comfort Control Fear Self-worth
	Health action process approach (Schwarzer, 1999)	Behaviors have a motivation/intention phase (affected by self-efficacy, outcome expectancies, and risk perception/threat) and a volition phase (subdivided into planning how to act, and action and maintenance phase which require effort, cognitive control, and self-efficacy).	Goal-directed	Control Fear
	Information–motivation–behavior skills (IMB) model (Fisher, Fisher, Bryan, & Misovich, 2002)	The IMB model is predominantly an intervention method assuming that people with high levels of information, motivation, and behavioral skills will undertake preventative health behavior.	Goal-directed	Fear Belonging

Attitude Theories mostly focus on the perceived cost and benefits of the outcome states and their contingencies (transitions). Attitudes usually underlie the theoretical basis of many theories of behavior and the main implication for interventions is that information or education should be used to provoke reflective change in such cognitions. Table 1 outlines the most prominent attitude models, which predominantly assume cognitive reflective processes that lead to the formation to various types of explicit beliefs. We also outline more recent, dual-process models that in addition assume automatic (habitual and innate) processes influencing attitude formation and change. We provide this taxonomy as a way of organizing the common aspects of the different models and their relation to our framework. It is evident from Table 1 that the various attitude theories assume mostly goal-directed computations, as defined in terms of our decision-theoretic (computational) approach to behavior change: anticipated states (e.g., financial, physical, or social rewards) characterized by their values, the transitions (probabilities) linking those states and actions, and the planning search through decision trees or mental models representing those quantities.

Goal Theories assume that behavior is not so much driven by immediate stimulation from the environment, but guided by intentional or desired states representing short-term and long-term goals. The goal construct is central to theories of self-regulation (see Austin & Vancouver, 1996, for definitions and reviews of the literature) and hence changing goals is assumed to result in behavior change. Goal theories usually assume constructing a model of the hierarchical structure of one's goals and hence the emphasis is on "planning" computations—constructing a decision tree with proximal and distal states and transitions. This process is computed in the brain by neural prediction error signals (indicating the difference between expected and actual states/rewards), which is formally known as model-based reinforcement learning (see Gläscher et al., 2010).

Stage Theories integrate aspects of both attitude and goal theories, acknowledging that behavior change unfolds in specific temporal steps—usually computations of values and transitions (probabilities of outcomes) precede the choice of a specific course of action or the formation of a behavioral goal. The dynamic nature of this cognitive process is assumed in the so-called stages-of-change models. It is evident that action planning appears during the later stages, although value computations continue throughout, as according to those theories, staying engaged in the behavior change process is determined by a "decisional balance"—a calculation of the pros and cons, both to the self and others, of trying to change (e.g., when the pros of preparation outweigh the cons from the contemplation stage). Stages-of-change models (e.g., health action process approach), similarly to some attitude and goal theories (e.g., theory of planned behavior, social–cognitive theory), stress the importance of environmental barriers and contingencies in being able to actually change and maintain behavior, known as *self-efficacy*. In reinforcement learning terms, self-efficacy is computed during planning as the "joint" probability of transition across states that represent stages of action completion (behavioral goals), which may also be reinforced by a specific reward—"feeling of control." In summary, interventions based on stage theories predominantly assume the involvement of goal-directed actions.

A Reinforcement Learning Perspective on Behavior Change Techniques

Progress in developing effective interventions requires understanding of how interventions work, that is, the mechanisms by which interventions cause behavior change (Albarracín et al., 2005; Michie & Abraham, 2004). The purpose of this section is to illustrate how ACT describes the mechanistic processes underpinning the most effective behavior change techniques derived from systematic reviews of the literature. Here we also discuss examples of some of those techniques in practice, which aims to illustrate the explanatory utility of the mechanistic action control architecture proposed here. Table 2 presents how each technique maps onto the behavioral determinants postulated in ACT—brain systems that generate unique psychological processes (including primary rewards or motivational states), which explain the workings of techniques proposed by three major frameworks: the *taxonomy of behavior change techniques* used in interventions (Abraham & Michie, 2008), *nudge theory* (Thaler & Sunstein, 2008), and the *mindspace* framework (Dolan et al., 2010, 2012a, 2012b). Nudge theory and the mindspace framework are based on insights from *behavioral economics*—a new discipline that seeks to combine lessons from psychology with those from economics, which provides the contrasting model of influencing behavior by using the more automatic processes of judgment and decision making (Kahneman, 2003; Kahneman & Tversky, 2000). Behavioral economics provides an account of how people actually respond to the context within which their decisions are made—the *choice architecture* (Thaler & Sunstein, 2008), which is an influential approach among the policymakers in the United States and United Kingdom. Table 2 presents the six principles (or nudges) of good "choice architecture." Mindspace is a mnemonic representing an elaborated and extended version of the Nudge framework, which outlines the nine most powerful influences on automatic behavior.[3] Next we discuss how ACT provides a mechanistic account of the techniques outlined in Table 2.

Influencing Goal-Directed Actions

Abraham and Michie (2008) present the most comprehensive classification of behavior change techniques used in interventions targeting health behaviors (see Michie et al., 2011, for an extended taxonomy of techniques). Table 2 also contains references to the behavioral theories (discussed in the previous section) that are represented or assumed by each technique. We classify those techniques according to the underlying computations (summarized in Figure 1) that they use.

Action-outcome contingency. According to ACT, information processed in the cognitive system is included in the calculation of the transition likelihoods between goal-directed actions and outcome states (see also Bandura, 1977, 1989). The typical techniques that target this process are *provide information about behavior-health link* and *provide information on consequences*. For example, a typical communication of health risk contains a proposition stating the perceived cause of a threat and a related effective coping action: "IF high blood pressure is caused by being unfit THEN exercise will reduce it" (Marteau & Weinman, 2006). Researchers and interventionists also try to uncover the characteristics of risk information that are likely to motivate behavior change (e.g., information about *actions* such as unhealthy behaviors, or *states* such as DNA risk regarding an inherited predisposition, which are likely lead to certain *values* such as various diseases). Such interventions also identify the cognitions to target (e.g., representations of threat)

[3] The MINDSPACE framework is becoming widely used across United Kingdom government, largely as a result of its use by the United Kingdom's Behavioural Insight Team in the United Kingdom Cabinet Office and Prime Minister's office—created to envision ways of supplementing the more traditional tools of government, with policies encouraging behavior change.

Table 2
Mapping Behavior Change Techniques in the Literature Onto the Classification Variables in ACT That Explain Their Mechanism of Action (The Techniques Are Grouped According to Publication)

			Classification variables in ACT	
Publication	Behavior change technique	Definition	Brain system	Computation
Taxonomy of behavior change techniques used in interventions (Abraham & Michie, 2008)	Provide information about behavior-health link (IMB)	General information about behavioral risk, for example, susceptibility to poor health outcomes or mortality risk in relation to the behavior.	Goal-directed	Outcome value Action-outcome contingency
	Provide information on consequences (TRA, TPB, SCogT, IMB)	Information about the benefits and costs of action or inaction, focusing on what will happen if the person does or does not perform the behavior.	Goal-directed	Outcome value Action-outcome contingency
Related theories: IMB = information-motivation-behavioral skills model; TRA = theory of reasoned action; TPB = theory of planned behavior; SCogT = social-cognitive theory; CT = control theory; OC = operant conditioning	Provide information about others' approval (TRA, TPB, IMB)	Information about what others think about the person's behavior and whether others will approve or disapprove of any proposed behavior change.	Goal-directed Reward	Outcome value Belonging
	Prompt intention formation (TRA, TPB, SCogT, IMB)	Encouraging the person to decide to act or set a general goal, for example, to make a behavioral resolution such as "I will take more exercise next week".	Goal-directed	Planning
	Prompt barrier identification (SCogT)	Identify barriers to performing the behavior and plan ways of overcoming them.	Goal-directed	Action-outcome contingencies planning
	Provide general encouragement (SCogT)	Praising or rewarding the person for effort or performance without this being contingent on specified behaviors or standards of performance.	Habit Reward	Action value Self-enhancement
	Set graded tasks (SCogT)	Set easy tasks, and increase difficulty until target behavior is performed.	Goal-directed Reward	Action-outcome contingencies planning Control
	Provide instruction (SCogT)	Telling the person how to perform a behavior and/or preparatory behaviors.	Goal-directed	Planning
	Model or demonstrate the behavior (SCogT)	An expert shows the person how to correctly perform a behavior, for example, in class or on video.	Goal-directed Reward	Planning Understanding Trusting
	Prompt specific goal setting (CT)	Involves detailed planning of what the person will do, including a definition of the behavior specifying frequency, intensity, or duration and specification of at least one context, that is, where, when, how, or with whom.	Goal-directed	Planning
	Prompt review of behavioral goals (CT)	Review and/or reconsideration of previously set goals or intentions.	Goal-directed	Outcome value planning
	Prompt self-monitoring of behavior (CT)	The person is asked to keep a record of specified behavior(s) (e.g., in a diary).	Goal-directed Reward	Action-outcome contingencies Control
	Provide feedback on performance (CT)	Providing data about recorded behavior or evaluating performance in relation to a set standard or others' performance, (i.e., the person received feedback on their behavior).	Goal-directed Reward	Action-outcome contingency Control

Table 2 (*continued*)

Publication	Behavior change technique	Definition	Classification variables in ACT	
			Brain system	Computation
	Provide contingent rewards (OC)	Praise, encouragement, or material rewards that are explicitly linked to the achievement of specified behaviors.	Habit Reward	Action value Greed Self-enhancement
	Teach to use prompts or cues (OC)	Teach the person to identify environmental cues that can be used to remind them to perform a behavior, including times of day or elements of contexts.	Habit	State-action contingency
	Agree on behavioral contract (OC)	Agreement (e.g., signing) of a contract specifying behavior to be performed so that there is a written record of the person's resolution witnessed by another.	Innate Reward	Approach Status
	Prompt practice (OC)	Prompt the person to rehearse and repeat the behavior or preparatory behaviors.	Habit	State-action contingency
	Use follow-up prompts	Contacting the person again after the main part of the intervention is complete.	Habit	State-action contingency
	Provide opportunities for social comparison (SCogT)	Facilitate observation of nonexpert others' performance for example, in a group class or using video or case study.	Habit Reward	Action value Belonging
	Plan social support or social change (social support theories)	Prompting consideration of how others could change their behavior to offer the person help or (instrumental) social support, including "buddy" systems and/or providing social support.	Goal-directed Reward	Planning Belonging
	Prompt identification as a role model	Indicating how the person may be an example to others and influence their behavior or provide an opportunity for the person to set a good example.	Goal-directed Reward	Outcome value Status Nurture
	Prompt self-talk	Encourage use of self-instruction and self-encouragement (aloud or silently) to support action.	Goal-directed Reward	Action-outcome contingency Self-enhancement
	Relapse prevention (relapse prevention therapy)	Following initial change, help identify situations likely to result in readopting risk behaviors or failure to maintain new behaviors and help the person plan to avoid or manage these situations.	Goal-directed Reward	Action-outcome contingencies Planning Control
	Stress management (stress theories)	May involve a variety of specific techniques (e.g., progressive relaxation) that do not target the behavior but seek to reduce anxiety and stress.	Reward	Comfort
	Motivational interviewing	Prompting the person to provide self-motivating statements and evaluations of their own behavior to minimize resistance to change.	Goal-directed Reward	Action-outcome contingency Outcome value Self-enhancement
	Time management	Helping the person make time for the behavior (e.g., to fit it into a daily schedule).	Goal-directed	Planning

(*table continues*)

Table 2 (*continued*)

Publication	Behavior change technique	Definition	Classification variables in ACT: Brain system	Classification variables in ACT: Computation
Nudge theory (Thaler & Sunstein, 2008)	Incentives	Incentives can influence decision making especially when they are salient for users.	Goal-directed Reward	Outcome value Greed
	Understand mapping from choice to welfare	Mental models aid users in their interactions with choice tasks. Transform information about possible outcomes associated with choice options into units that translate easily into actual use.	Goal-directed	Planning
	Defaults	Options that are assumed as preselected if the individual does not make an active choice of another available alternative.	Innate Reward	Avoidance Comfort
	Give feedback	Salient warning signs in a way that gives information when people are doing well and when they are making errors.	Habit	State-action contingency
	Expect error	Assume error is inevitable and make the required action a habit using recurrent cues and prompts.	Habit	State-action contingency
	Structure complex choices	Redesigning the choice environment when choosing between complex multi-attribute alternatives (e.g., breaking it down into easier chunks).	Habit	State-action contingency
MINDSPACE framework (Dolan et al., 2010, 2012a, 2012b)	Messenger	We are heavily influenced by who communicates information to us.	Habit Reward	Action value Trusting
	Incentives	Our responses to incentives are shaped by mental states such as greed and fear of losses.	Innate Reward	Approach Avoidance Greed Fear
	Norms	We are strongly influenced by what others do.	Habit Reward	Action value Belonging
	Defaults	We 'go with the flow' of preset options.	Innate Reward	Avoidance Comfort
	Salience	Our attention is drawn to what is novel and seems relevant to us.	Habit	State-action contingency
	Priming	Our acts are often influenced by subconscious cues.	Habit	State-action contingency
	Affect	Our emotional associations can powerfully shape our actions.	Innate Reward	Approach Avoidance Fear Attraction
	Commitments	We seek to be consistent with our public promises, and reciprocate acts.	Habit Reward	Action value Status
	Ego	We act in ways that make us feel better about ourselves.	Innate Reward	Approach Avoidance Self-enhancement

so as to optimize the motivational impact of risk information (see Marteau & Weinman, 2006). Marteau and Lerman (2001) also demonstrate that providing information about genetic risk, and thus, creating mental associations between states and outcome values (adverse consequences), may not be sufficient to increase motivation to change behavior; but that change is more likely if people are persuaded that changing their behavior can indeed reduce the risk of adverse health outcomes (i.e., the perceived action-outcome contingency changes too). Similarly, Ruiter, Kok, Verplanken, and Brug (2001) show that motivating protective actions requires that the recommended action is judged as effective and feasible.

Outcome value. It is evident that many traditional behavior change techniques target the goal-directed system by *providing information about behavior-health link* and *providing information on consequences* (in the form of messages, education, advice), which, assuming human rational nature, is supposed to persuade and/or train recipients to adopt a specific behavior.

Persuasion messages may contain arguments describing the benefits of adopting a healthy behavior (e.g., physical wellness) and/or costs of unhealthy behaviors (e.g., heart disease; Gray, 2008). Note that in such interventions the goal-directed system receives inputs from the reward system too (see Figure 1). For example, respondents provoked to experience high (vs. low) fear of breast cancer, by manipulating the information about severity of the disease, are more likely to base their attitude to breast self-examination on the persuasive arguments (Ruiter et al., 2001). Likewise, persuasive arguments are more likely to increase intentions to engage in physical exercise when negative affect is evoked by a background picture showing an overweight woman in distress (McCormick & McElroy, 2009). Persuasion techniques may also include *providing information about others' approval*. Such arguments aim to change the perceived value of the expected state outcomes, and also aim to suggest different goal-directed pathways to achieve and avoid those outcomes (planning).

Another set of techniques labeled as *incentives* is used to directly influence the values of outcome states (e.g., even when the ultimate outcome of the actions is different such as better health). Specifically, people are sensitive to prices and costs, which is known as the economic *law of demand* (Kreps, 1990; Pearce, 1986). For example, higher cost reduces consumption of alcohol and cigarettes (Chaloupka, Grossman, & Saffer, 2002; Colman, Grossman, & Joyce, 2003; Manning, Blumberg, & Moulton, 1995); and as constraints on drug use decrease (e.g., drugs are easily available at low cost, or there is little in the way of alternatives to be forfeited by using drugs) consumption increases (Chaloupka, Grossman, Bickel, & Saffer, 1999). People also impose on themselves penalties for failing to act according to their long-term goals (Trope & Fishbach, 2000), such as when students are willing to self-impose costly deadlines to overcome procrastination (Ariely & Wertenbroch, 2002). Note that goal-directed incentives-based interventions also involve introducing rules-and-sanctions, in the form of legislation and regulation (e.g., smoking ban, compulsory seatbelts), which are also effective way to change the values of outcomes and consequently behaviors (Chaloupka & Grossman, 1996; Chaloupka, Grossman, & Saffer, 2002).

Planning. People need to continuously maintain knowledge of the decision tree (including its short-term and long-term states, actions and values), which enables planning, or finding a good policy, and thus, guides their behavior change. This process is triggered by techniques such as *plan social support or social change* and *relapse prevention*. For example, a successful case of relapse prevention is when a Philippine bank offered a saving product intended for individuals who want to restrict access to their savings (Ashraf, Karlan, & Yin, 2006), which was very popular among Philippine women who are traditionally responsible for household finances and need solutions to temptation problems.

Another technique using this principle is *prompt intention formation*. Of interest to the authors, people are also more likely to change behavior (e.g., quitting smoking) when they generate more avoidance goals or outcomes (e.g., getting rid of hacking cough, not developing cancer and heart disease, get rid of smell of smoke on clothes and belongings; see Worth et al., 2005). In addition to long-term goals (desired outcomes), it is also essential to be able to generate short-terms goals that are vivid and detectable, which allow people to monitor their progress (e.g., get rid of the cough; Rothman, Hertel, Baldwin, & Bartels, 2007). Planning techniques such as *prompt specific goal setting*, *prompt reviews of behavioral goals* are used to enable this process.

Some effective techniques, such as *prompt specific goal setting*, aim to stimulate goal-directed as well as habitual actions. This technique is also known as forming *implementation intentions* (Gollwitzer, 1993, 1999; Gollwitzer & Brandstatter, 1997), which are plans that specify when, where, and what behavior will be performed (e.g., *during my lunch break, I will visit the canteen and eat one salad and two fruits*). Implementation intentions promote goal attainment by designing specific plans for action that usually take the form of if-then rules (if state *x*, then action *y*; Gollwitzer, 1993; Sheeran, Webb, & Gollwitzer, 2005). This technique improves self-regulation by eventually relying on situational cues that elicit behavior automatically, often without conscious decision making (e.g., acting after some regular activity, such as taking a tablet after teeth brushing); which is similar to development of behavioral habits (Lally et al., 2010). The effectiveness of the implementation intentions technique is consistent with recent neural and computational evidence suggesting goal-directed control is involved in training model-free, habitual responses (see Dolan & Dayan, 2013).

Understand mapping from choice to welfare is also a goal-directed technique, which helps people to understand the mapping from action to state values. This involves transforming information about possible outcomes (e.g., health states) associated with available choice options (e.g., medical treatments) into units that translate more easily into actual use for planning. Thaler and Sunstein (2008) illustrate this technique by pointing out that when buying apple to make apple cider, it helps to know that three apples make one glass of cider.

Influencing Habitual Actions

The distinguishing feature of habit-related techniques is the requirement for repetition to develop automatic behaviors. Different techniques outlined in Table 2 tend to focus on influencing how individuals attribute values to actions and how individuals compute the contingency or association between states and actions. We group them accordingly under each element.

State-action contingency. Techniques such as *prompt practice*, *use follow-up prompts*, *give feedback*, *expect error*, *salience*, and *priming*, are designed to facilitate mental processing of state cues and improve their association with actions, which should maximize the formation and triggering of habitual behavior.

Give feedback is a good example of a technique promoting habit development, which involves providing warning cues when people are doing well and when they are making errors, which is akin to formal models of habit learning (see Dayan & Niv, 2008).

Expect error is also a way to utilize the habit system by providing recurrent cues to prompt motor habits (e.g., making pill intake habitual by either taking placebo pills for the days without a pill, or by taking the pill after some regular daily activity).

Structure complex choices is useful when deciding between multiattribute alternatives (e.g., ordering choice alternatives according to the most important cue, or, according to their similarity). This technique is usually used either to make choice environments manageable by mental habits, such as *lexicographic heuristics* (deciding on the basis of one cue at a time) and automatic search for salient cues for similarity (see Mussweiler, 2003), or to enable goal-directed planning in complex decision trees. Such restructuring (translation) in terms of a single most useful attribute complies with mental habits such as judgment heuristics that operate with one-reason-at-a-time (Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Goldstein, 2008). This intervention technique (focusing recipients' attention on the most important choice attribute) is illustrated in a field study of how information on HIV risk changes sexual behaviors among teenagers in Kenya (Dupas, 2011). Providing information on the relative risk of HIV infection by partner's age group led to a 28% decrease in teen pregnancy and 61% decrease in the incidence of pregnancies with older, riskier partners (in contrast, the national HIV education curriculum achieved no significant results). By making the age of partner salient, the intervention transformed a complex multiattribute choice into a habitual action triggered by a single state cue.

Salience is another important technique, because habits are triggered by salient state cues. Orbell and Verplanken (2010) report how salient cues elicit wanted and unwanted habitual responses (e.g., smoking when drinking alcohol in a pub, and dental flossing in response to a specified situational cue). Habitual consumption is triggered by salient positioning of healthy foods at the beginning of the queue in canteens, whereas less healthy foods are positioned last at the least visible places (see Thaler & Sunstein, 2008, pp. 1–3). Conversely, salient stimuli can also increase the likelihood of succumbing to unhealthy habits, such as when the probability of purchasing an extra item at a supermarket checkout increases with the time spent waiting (Houser, Reiley, & Urbancic, 2008). Lally, Chipperfield, and Wardle (2008) successfully used this approach in an intervention to enable the participants to control their weight, which was delivered as a leaflet containing advice on habit formation—simple recommendations such as eating roughly at the same salient times and incorporating the target behaviors into salient daily routines. Lally et al. (2010) extend this approach to developing healthy habit formation across variety of healthy eating, drinking, and activity behaviors.

Priming techniques utilize existing state-action associations and trigger habitual actions by making the habitual actions more accessible in memory, which often happens with less conscious processing. For example, interventions that alter subtle cues in eating situations have been shown to control eating habits. Sobal and Wansink (2007) demonstrated that the amounts of food and drink that people serve and consume decrease with smaller sizes of plates, spoons, and glasses. Smaller plates lead to lesser food intake, because people habitually consume around 92% of what they serve themselves (Wansink & van Ittersum, 2011), that is, habitual control of food intake is cued by the amount of food on a plate (Wansink & Cheney, 2005). TV advertising of healthy food also primes consumption of unhealthy snacks, which is not related to reported hunger or conscious influences (Harris, Bargh, & Brownell, 2009). Physical sensations like smells can also prime habits: mere exposure to scent of an all-purpose cleaner prompted people to keep their table clean while eating in a canteen (Holland, Hendriks, & Aarts, 2005). Similar subtle effects of cues were used to prime healthy lifestyle habits by asking participants to make a sentence out of scrambled words such as *fit*, *lean*, *active*, *athletic*, which resulted in more participants using the stairs instead of elevators (Wryobeck & Chen, 2003). In another domain, Aarts and Dijksterhuis (2003) demonstrate likewise that people habitually speak quietly when cued by pictures of a library, which could be a cost-effective way to control noise in public places. Habitual actions can also be primed by simply measuring intentions, because such questions alter the ease of cueing habits; for example, asking people to indicate the likelihood of flossing their teeth in the coming week increases the frequency of this behavior over that period (Levav & Fitzsimons, 2006).

Action value. Techniques such as providing *contingent rewards*, *messenger*, *norms*, and *commitments* usually aim to motivate individuals to select appropriate (mental or physical) action routines. *Providing contingent rewards* is used in interventions developing constructive habits (see Wolpe, 1990). This reinforcement principle has been successfully used to treat drug addiction and substance misuse (including smoking and alcohol consumption) and to improve medication compliance (Higgins, Heil, & Lussier, 2004). Such interventions usually include earning points contingent on patients submitting drug/substance-negative urine specimens. Points usually begin at a low value (e.g., $2.50) and increase with each consecutive negative test result (points can then be exchanged for retail items kept onsite at the clinic or clinic staff makes all purchases). A drug-positive result or failure to provide a scheduled specimen reset the voucher value back to the initial low value from which it could escalate again (Roll, Higgins, & Badger, 1996). Providing such contingent rewards can also reinforce people to lose weight and take more exercise (Paul-Ebhohimhen & Avenell, 2008; Volpp et al., 2008), eat healthier foods, drink less alcohol, and give up smoking (Marteau, Ashcroft, & Oliver, 2009; Sindelar, 2008).

Messengers are usually physically attractive, or have high social status and authority. Messenger effects are reported in a meta-analysis of 166 HIV-prevention interventions, which found that the demographic and behavioral similarity between the interventionist and the recipients facilitates behavior change (Durantini et al., 2006). Recipients also trust persuasive messages from a perceived authority (Cialdini, 2007). Using messengers is a classic persuasion technique understood to operate as a mental heuristic (habit), such as "if he is suggesting it, therefore it must be good" that is, high value. This technique is classified as affecting the computation of action value in habits.

Norms can also trigger actions in automatic ways (Aarts & Dijksterhuis, 2003; Cialdini & Goldstein, 2004; Cialdini, Kallgren, & Reno, 1991). Using norms to motivate behavior change is extensively reported in the literature. For example, Goldstein, Cialdini, and Griskevicius (2008) demonstrated that a hotel-towel reuse signs conveying information about social norms (that most guests at the hotel recycled their towels, or that most previous occupants of the room had reused towels) is more persuasive than a widely adopted sign containing the basic environmental-protection message asking guests to help to save the environment (see Cialdini, 2003, for related interventions aiming to reduce antisocial behavior). Schultz et al. (2007) also used social norms to motivate environmental behaviors—messages describing average energy usage in the neighborhood produced energy savings when households were already consuming at a high rate.[4] Social norms have also proven very effective in interventions motivating preventative health behaviors such as hand washing (Curtis, Danquah, & Aunger, 2009; Judah et al., 2009; Perkins, 2004; Whitby et al., 2006). Ybarra and Trafimow (1998) demonstrated that provoking feeling of belonging—by asking respondents what they have in common with their family and friends—makes subjective norms the strongest predictor of intentions toward using a condom during sex. Similarly, men reported increased motivation to undergo cancer screening when given information indicating higher rates of participation in screening among men (Sieverding, Decker, & Zimmermann, 2010). Conformity with group norms has also been elicited in interventions to reduce alcohol use in American student-athletes (Perkins & Craig, 2006). To summarize such evidence, a meta-analysis of 21 studies (Rivis & Sheeran, 2004) found a medium to strong correlation between descriptive norms and intention for health-risk behaviors (smoking, drug use, binge drinking, condom use, extradyadic sex, and gambling) and health-promoting behaviors (healthy eating, dieting, and physical exercise).

The key question is how social norms induce those automatic behavior changes. Psychologically, people automatically use peer norms as a standard against which to compare their own behaviors (Clapp & McDonell, 2000; Perkins & Berkowitz, 1986) and adhering to norms triggers rewarding feeling of *belonging* (maybe because the social group is the environment to which humans adapt to survive, see Fiske, 2010). Crucially, recent evidence reveals that those processes are reinforced by dopamine reward networks in the brain. Klucharev et al. (2009) found that conflict with group opinion activates a neuronal response in the rostral cingulate cortex (related to monitoring response conflicts/errors and in differential processing of aversive outcomes) and also deactivates the ventral striatum (nucleus accumbens—representing the value of the expected reward and decreases for aversive stimuli). The amplitude of this signal change (especially in the ventral striatum) correlated with differences in conforming behavioral adjustments. This effect is similar to the "prediction error" signal suggested by models of reinforcement learning, which provides evidence that social norms cause conformity through learning mechanisms. This finding implies that habit-based action value learning is the most plausible underlying mechanism that mediates the effect of norms on behavior change. In other words, available actions closer to the norm are automatically valued as less aversive and more rewarding (e.g., feeling of belonging resulting conforming to the norm); and vice versa for actions conflicting with the norm. This automatic valuation process may also be driven by mental heuristics such as "if majority are doing it, then it must be a good action"—such mental habits may be acquired/reinforced during evolution and individual socialization (see Erev & Barron, 2005).[5]

Commitments are techniques that involve asking people to make a public (verbal or written) promise to execute a specific action or to achieve a specific goal. As a result, individuals automatically tend to be consistent with their public promises, which is associated with appetitive (gaining status) or aversive (loosing status) outcomes, respectively. Commitments to action are usually more impactful when individuals are asked to write the commitment down (Cioffi & Garner, 1996). For example, a randomized controlled trial showed that Black women signing a behavioral contract were more likely to reach their exercise goals than a control group where no commitment was made (Williams, Bezner, Chesbro, & Leavitt, 2006). In an attempt to reduce the number of patients who do not attend their appointments (DNAs), patients were asked to write down the time and date on the appointment card themselves (rather than the common practice of the nurse doing so), which resulted in a reduction in subsequent DNAs of 18% (Martin, Bassi, & Dunbar-Rees, 2012). In another intervention involving voluntary commitment on recycling (i.e., influencing environmental behavior change) participants were more likely to enroll in a curb-side recycling program if they had made a written commitment to do so than if they are just informed about the program (Werner et al., 1995). Commitments, we believe, are a type of mental heuristics (habits) such as "I must act [or achieve this goal] because I promised so." Thus, the public promise is automatically used as a reason for the action, that is, inferring that the action is a "good" (valuable) thing to do, instead of justifying (valuing) the action on the basis of its consequences. Such heuristics are either learned (reinforced) during early years of socialization, or they may have evolved because of their adaptive role in society. Note that the commitment mechanism increases the value of an action if the contract is defined in terms of what the person should do (e.g., regular visits to the gym), but the mechanism may also increase the value of an outcome if the contract is defined in terms of what the person will achieve (e.g., losing weight). Future research using neuroimaging should answer this question Furthermore, because of commitments' reward-related properties, they may also increase the association strength in memory between the state cues and the action; thus, improving state-action contingency. Overall, commitments work in complex ways and the answer may depend on the specific intervention circumstances and commitment technique.

Influencing Innate Actions

Innate actions are based on a representation of the contingency between the state and the reward outcome (state-state associations). Those computations instigate two classes of evolved, automatic reactions broadly defined as approach and avoidance. Recall the key defining feature of innate actions is their automaticity, especially in one-off situations, regardless of whether or not the action leads to reward in the current situation; and thus, innate actions differ from habitual actions in which behavior is strengthened or weakened by its consequences (so repetition is required). Different techniques outlined in Table 2 have been used in inter-

[4] Keizer, Lindenberg, and Steg (2008) demonstrate normative effect on behavior in the opposite direction—showing that criminal activity can be made more likely, through observation of others' behavior. In particular, six field experiments demonstrate that graffiti or littering can encourage stealing and violation of police ordinances in various social contexts. The authors conclude that "as a certain norm-violating behaviour becomes more common, it will negatively influence conformity to other norms and rules" (p. 1684).

[5] Note that we cannot make inferences about the impact of norms on goal-directed actions, because, so far, there is no evidence for impact upon the value of the instrumental outcome (e.g., energy savings). Similarly, social norms are not characterized as operating in the service of meeting the goal of attaining belonging, because this implies that such (mostly unconscious) goal-directed action substitutes the conscious goal-directed task (e.g., related to the energy saving). We are not aware of empirical evidence supporting this assumption.

ventions to facilitate either approach or avoidance actions, which is how we classify those examples here.

Approach. Techniques such as incentives, *norms*, *affect*, and *ego*, have been used to trigger approach actions in response to information about states that predict positive rewards.

Incentives have been used to highlight the immediate personal benefits of behaviors, which triggers automatic one-off approach actions. In a field study conducted in Malawi, Thornton (2008) used economic incentives to get people to pick up their HIV result. The biggest increase in uptake (by 50%) was caused by the change from zero incentive to a very small payoff (10–20 cents or one-tenth of a day's wage), while offering more money further affected behavior to a much lesser degree. In Pavlovian terms, offering some reward commands an impulsive approach to its predictor (the HIV test center). Similar effects of small incentives is observed in field study in Kenya, which showed that small time-limited reductions in the cost of purchasing fertilizer at the time of harvest induce substantial increases in fertilizer purchase (i.e., farmers snatch immediate small payoffs), as much as considerably larger price cuts later in the season (Duflo, Kremer, & Robinson, 2010).

Affect techniques have been used in interventions that provoke feelings that directly motivate automatic approach responses. Bertrand, Karlan, Mullainathan, Shafir, and Zinman (2010) cued impulsive purchase in a field experiment in South Africa, which varied the price and creative content of loan offers made to former clients of a subprime consumer lender. In particular, beautiful smiling female photo on the letter describing the loan offer significantly increased demand by as much as a 25% reduction in the loan's interest rate. The photo is irrelevant to the expected financial outcomes in this instrumental task. However, the photo creates an approach action toward the associated state (the loan offer), and thus, biases decision making, because the latter becomes a predictor of positive reward (feeling of attraction).

Ego techniques use state cues associated with rewarding outcomes such as positive self-image and feeling of self-enhancement (self-worth), which instigate automatic approach actions, partly because humans are attuned to cues indicating reputational consequences of behavior (Haley & Fessler, 2005). For example, introducing school uniforms (cues for social status and self-worth) in a large urban school district improved attendance rates, as students are more likely to "approach" the school when it became a "state" predictive of social status (Gentile & Imberman, 2009). In another intervention, exposure to favorable ego characteristics of people who exercise (e.g., appearance, health, energy, achievements, or relationships) increased exercise participation (Ouellette et al., 2005). In this case, the gym is "approached" because it becomes a cue (conditioned stimulus) associated with rewarding status-related outcomes (unconditioned stimuli). In addition, increased vigor of exercise is achieved via the general Pavlovian-Instrumental transfer mechanism when people are already habitually exercising.

Avoidance. Techniques such as *defaults*, *affect*, and *ego*, have been used to motivate avoidance actions in response to information about states predicting aversive outcomes. *Defaults* have been mostly used in interventions changing financial behaviors, when the default is to automatically enroll employees in their pension plan, about three-quarters tend to retain both the default contribution rate and the default asset allocation; also, introducing a participation default can increase participation rates among new employees by more than 50% (Choi, Laibson, Madrian, & Metrick, 2002, 2003; Madrian & Shea, 2001; Thaler & Benartzi, 2004). Such powerful effects of defaults on behavior have been observed in a wide range of other settings like organ donation decisions (Abadie & Gay, 2006; Johnson & Goldstein, 2003), choice of car insurance plan (Johnson, Hershey, Meszaros, & Kunreuther, 1993), car option purchases (Park, Jun, & MacInnis, 2000), consent to receive e-mail marketing (Johnson, Bellman, & Lohse, 2002), employees' contributions to health care flexible-spending accounts (Schweitzer, Hershey, & Asch, 1996), and vaccination and HIV testing for patients and health care workers (Halpern, Ubel, & Asch, 2007). Such effects are most likely caused by automatic avoidance of actions that result in aversive outcomes such as immediate opting-out costs or effort (made salient by human tendency to overvalue the present), which leads to procrastination (Johnson & Goldstein, 2003; Laibson, 1997; O'Donoghue & Rabin, 1999; Samuelson & Zeckhauser, 1988).

Affect is also used in interventions using states that provoke feelings that directly motivate automatic avoidance responses. A good example is the British Heart Foundation's, 2004, "Give up before you clog up" campaign (National Social Marketing Foundation, 2006), which was based around revolting, disgust-provoking images of cigarettes filled with glutinous fat deposits and extruding viscous white fat rather than ash during smoking. The British Heart Foundation made this aim explicit when commenting on a poster (showing a cigarette as a fat-filled artery) that they wished to "develop a Pavlovian response between the cigarette and the gunk-filled artery so that as soon as a smoker sees a cigarette they will be reminded of the clogged artery" (Hawkes, 2004). As a result, quit attempts among NHS Stop Smoking Services increased significantly (almost doubled) in January–March, 2004, the phase when the TV advert was featured heavily. Wakefield, Loken, and Hornik (2010) report that similar TV campaign resulted in a decline in adult smoking rates in Australia. Those campaigns illustrate outcome-specific (goal-directed) Pavlovian-instrumental transfer (PIT) involving a Pavlovian cue (cigarette) associated with an aversive outcome (disgust) that is also related to a specific action (smoking), which inhibits the action (for very habitual, automatic smokers this intervention could work as a general PIT reducing the vigor of smoking).

Similar outcome-specific PIT is observed when cues such as words associated with an aversive outcome, provoking fear, suppress an instrumental action predicting the same outcome (De Martino et al., 2006; Kahneman & Frederick, 2007; Seymour & Dolan, 2008). In particular, *loss*-framed persuasive messages emphasizing the dangers of not obtaining mammography have stronger impact on opting for mammography than *gain*-framed messages emphasizing the benefits of obtaining mammography, even though both messages are factually equivalent (Banks et al., 1995). Loss-framed messages also elicit greater behavioral intention for breast and cancer self-examination (McCormick & McElroy, 2009; Webb & Sheeran, 2006).[6] This suggests that when faced with the choice dilemma "screen" versus "not screen," the aversive cues repel the individual away from the latter option.

Affect techniques have also been used in a form of Pavlovian-learning where neutral states are conditioned to motivate more

[6] Recent meta-analyses report that this effect is not universal and some health behaviors are not significantly affected by gain-versus-loss frames (O'Keefe & Jensen, 2007, 2008).

complex avoidance actions. Curtis, Garbrah-Aidoo, and Scott (2007) discovered that only 3% mothers in Ghana wash hands with soap after toilet use, because Ghanaians use soap when they felt that their hands were dirty, usually after cooking or traveling, which also provokes feelings of disgust. Those findings informed an intervention campaign, TV commercials, focusing on associating toilets with dirt and emotional disgust response rather than promoting soap use. For example, soapy hand washing was shown only for 4 s in one 55-s video clip, but there was a clear message that toilets prompt worries of contamination—the central idea was that there is something invisible on hands and mothers should feel that their hands were contaminated. This campaign resulted in a 13% increase in the use of soap after the toilet and 41% increase in reported soap use before eating. Similarly, a short message provoking disgust has been shown to promote hand-washing with soap after using a public toilet in a Western society (Judah et al., 2009; Whitby et al., 2006). Seymour, Singer, and Dolan (2007, p. 303) outline the complex coordination of several processes involved in learning such avoidance actions: Pavlovian learning of predictors (e.g., toilets) of the aversive outcomes (e.g., contamination, disgust); Pavlovian and/or instrumental escape from these predictors (e.g., approaching sink and hand washing with soap); establishment of positive value of the aversive inhibitor—the safety state (e.g., clean hands); and instrumental reinforcement of the avoidance action by the safety state which may be dopamine-dependent (e.g., feeling of comfort).

Ego techniques have been used to provoke avoidance of harmful actions by presenting states or stimuli associated with aversive outcomes such as negative self-image. For example, smokers with negative images of smokers are more successful at quitting, that is, inhibiting or avoiding smoking (Gerrard et al., 2005). Other good examples include using unfavorable social images of characters engaging in risky behaviors to reduce willingness to engage in unprotected sex (Gibbons & Gerrard, 1995), drinking (Gerrard et al., 2002), and drink driving (Gibbons et al., 2002). Similarly, avoidance or suppression of harmful actions has been achieved after showing states illustrating the negative effects of sun exposure on physical appearance, such as ultraviolet photos of skin appearance or photo aging (premature wrinkling and age spots), which reduced harmful ultraviolet exposure from tanning booths (Gibbons et al., 2005; Jones & Leary, 1994) and sunbathing (Mahler et al., 2003). In a different real-life situation, avoidance or inhibition of socially harmful actions was achieved in an intervention that placed a poster of frowning eyes above an honesty box where people can get their drinks and pay suggested prices without being supervised, which resulted in less cheating and threefold increase in payment frequency (Bateson, Nettle, & Roberts, 2006).

Discussion and Conclusions

This article offers a novel, mechanistic approach to behavior change based on reinforcement learning theory, which is informed by recent advances in cognitive and behavioral neuroscience. In addition to popular descriptive "dualisms" assumed in dual-process models of behavior, ACT provides a mechanistic-level account of several competing neuropsychological systems for action control, which have not been distinguished in models of behavior change. The proposed account works at a different level of analysis and, therefore, promises to lead to the development of a more general and integrative theory of behavior change. Thus, the truly "novel" aspects of ACT theory is in the implications for analyzing how therapeutic behavioral interventions are best attained. We also describe the major behavior change theories and techniques in terms of these action systems and their interaction (i.e., this article is about analyzing how interventions work, not about how to generate them). Therefore, ACT provides a methodology for linking the means for motivating action (behavior change techniques) and the mechanisms for action. This is an important theoretical advance (see Michie et al., 2008), because it unifies in a single model the determinants of behavior (social, cognitive, or environmental) and the techniques to change these determinants (Sniehotta, 2009). ACT is also a general framework that enriches theories of behavior change—providing missing causal mechanisms, which can be used to refine theory on the basis of intervention evaluations.

Future Research

We recognize that there are at least three unresolved conceptual and empirical issues that require further research.

Deriving theoretical and practical predictions. This research should attempt to specify if-then conditionals that could be used to either test the theory or draw inferences about the relative effectiveness of different interventions and/or their effectiveness under different conditions. For example, the reward system is linked to several other concepts, but these links do not provide any if-then conditionals that could be used to derive a novel prediction or draw inferences about the relative effectiveness of different interventions or their effectiveness under different conditions. The same issue applies to the discussion of the three control systems and different intervention techniques. To a large extent, the current framework is able to offer a new classification and phenomenological analysis of existing theories and findings, which can be useful to identify links between existing theories and empirical findings, and to uncover unrecognized gaps in the literature. Future developments should aim to identify more theoretical and empirical implications.

In this respect, although many dual-process theories suffer from major conceptual problems, there are several theories that include specific assumptions about the relative effectiveness of different behavior change techniques. For example, several attitude theories explicitly discuss the relative effectiveness of different interventions to change spontaneous versus deliberate responses (e.g., Gawronski and Bodenhausen's (2006) *associative*—propositional evaluation model; McConnell and Rydell's (2014) *systems of evaluation model*; Petty and Briñol's (2014) *metacognitive model*). Similarly, Wood and Neal (2007) offered a detailed theoretical, psychological analysis of the habit-goal interface. With regard to intervention strategies that have been inspired by these theories, there is also a growing body of research in the clinical domain, comparing the relative effectiveness of interventions targeting higher-order processes (e.g., cognitive behavior therapy) versus lower-order processes (e.g., cognitive bias modification). Although these theories have a narrower focus on specific phenomena, many of them include detailed assumptions about the relative effectiveness of different interventions and their effectiveness under different conditions. The current version of ACT is only the first step in this direction and it needs further elaboration to offer more precise empirical predictions that could be used to test the theory or provide practical information on how to improve the effectiveness of interventions to change behavior.

Investigating the domains of application. The elements of ACT can be applied in isolation or in combinations, which should depend on what intervention is effective in the specific circumstances. However, triggering certain regulatory processes (e.g., habitual imitation) may be more effective for changing specific behaviors (e.g., dieting). Therefore, future research should determine the relative effectiveness, and also cost-effectiveness, of the various action systems in different domains (e.g., health, environment, or finance), behaviors (e.g., smoking, recycling, or saving), and populations (e.g., sociodemographic groups).

Sustainability of interventions targeting each control system. This issue relates to studying whether behavior change is short-term or long-lived depending on the targeted regulatory action system. Interventions relying on habitual and innate actions—triggered by contextual cues—are likely to be long-term, because recurring real-world environments (e.g., jobs, schools, neighborhoods, and culture in general) can provide frequent sources of contextual influence, without the need for computationally intensive goal-directed control. For example, an ingenious recent intervention succeeded in reducing driving speed on dangerous "S" curves along Chicago's Lake Shore Drive by painting a series of white stripes onto the road, which are initially evenly spaced but get closer together as drivers reach the most dangerous section of the curve (Thaler & Sunstein, 2008, p. 37). These contextual cues give the sensation that driving speed is increasing (even when the speed does not really change), because of the workings of automatic perceptual processes, which in turn triggers the driver's natural instinct or habit to slow down.

How to use ACT in intervention design. Finally, several frameworks for designing interventions have already been developed, which can be combined with ACT to design more theoretically grounded and effective interventions. A leading example in the health sphere is *intervention mapping* (IM; Bartholomew et al., 1998), which sets out a five-stage process for intervention development, running from setting objectives through to generating evaluation plans. The five steps are preceded by a needs-assessment-stage, which calls for precisely specifying the target behavior to be changed and explicitly requires the selection and adoption of relevant models from behavioral theory (see also Abraham & Michie, 2008; Darnton, 2008; Michie, van Stralen, & West, 2011). IM sets behavioral models at the center of the intervention planning process—in helping to understand specific behaviors and identifying the underlying factors that influence them. As a result, the intervention strategy followed through all the subsequent steps is shaped by the particular model selected at the beginning. This approach can easily integrate elements from ACT.

We conclude, therefore, that there are many unresolved questions and many exciting avenues for future research, and our framework provides a fresh perspective from which conceptual developments and empirical advances can be made.

References

Aarts, H., & Dijksterhuis, A. (2003). The silence of the library: Environment, situational norm, and social behavior. *Journal of Personality and Social Psychology, 84,* 18–28. http://dx.doi.org/10.1037/0022-3514.84.1.18

Abadie, A., & Gay, S. (2006). The impact of presumed consent legislation on cadaveric organ donation: A cross-country study. *Journal of Health Economics, 25,* 599–620. http://dx.doi.org/10.1016/j.jhealeco.2006.01.003

Abraham, C., & Michie, S. (2008). A taxonomy of behavior change techniques used in interventions. *Health Psychology, 27,* 379–387. http://dx.doi.org/10.1037/0278-6133.27.3.379

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50,* 179–211. http://dx.doi.org/10.1016/0749-5978(91)90020-T

Albarracín, D., Gillette, J. C., Earl, A. N., Glasman, L. R., Durantini, M. R., & Ho, M. H. (2005). A test of major assumptions about behavior change: A comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. *Psychological Bulletin, 131,* 856–897.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The anterior cingulate cortex. The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences, 935,* 107–117. http://dx.doi.org/10.1111/j.1749-6632.2001.tb03476.x

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111,* 1036–1060. http://dx.doi.org/10.1037/0033-295X.111.4.1036

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York, NY: HarperCollins.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science, 13,* 219–224. http://dx.doi.org/10.1111/1467-9280.00441

Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics, 121,* 635–672. http://dx.doi.org/10.1162/qjec.2006.121.2.635

Aunger, R., & Curtis, V. (2008). Kinds of behaviour. *Biology and Philosophy, 23,* 317–345. http://dx.doi.org/10.1007/s10539-007-9108-4

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin, 120,* 338–375. http://dx.doi.org/10.1037/0033-2909.120.3.338

Balleine, B. W. (2005). Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior, 86,* 717–730. http://dx.doi.org/10.1016/j.physbeh.2005.08.061

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35,* 48–69. http://dx.doi.org/10.1038/npp.2009.131

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191–215. http://dx.doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44,* 1175–1184. http://dx.doi.org/10.1037/0003-066X.44.9.1175

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes, 50,* 248–287. http://dx.doi.org/10.1016/0749-5978(91)90022-L

Banks, S. M., Salovey, P., Greener, S., Rothman, A. J., Moyer, A., Beauvais, J., & Epel, E. (1995). The effects of message framing on mammography utilization. *Health Psychology, 14,* 178–184. http://dx.doi.org/10.1037/0278-6133.14.2.178

Bargh, J. A. (1996). Automaticity in social psychology. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 169–183). New York, NY: Guilford Press.

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 457–481). New York, NY: Guilford Press.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479. http://dx.doi.org/10.1037/0003-066X.54.7.462

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology, 71,* 230–244. http://dx.doi.org/10.1037/0022-3514.71.2.230

Barr, N. A. (1987). *The economics of the welfare state*. London: Weidenfeld & Nicolson.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130,* 553–573. http://dx.doi.org/10.1037/0033-2909.130.4.553

Bartholomew, L. K., Parcel, G. S., & Kok, G. (1998). Intervention mapping: A process for developing theory- and evidence-based health education programs. *Health Education & Behavior, 25,* 545–563. http://dx.doi.org/10.1177/109019819802500502

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2,* 412–414. http://dx.doi.org/10.1098/rsbl.2006.0509

Berridge, K. C., & Kringelbach, M. L. (2008). Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology, 199,* 457–480. http://dx.doi.org/10.1007/s00213-008-1099-6

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology, 9,* 65–73. http://dx.doi.org/10.1016/j.coph.2008.12.014

Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., & Zinman, J. (2010). What's advertising content worth? Evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics, 125,* 263–305. http://dx.doi.org/10.1162/qjec.2010.125.1.263

Blair, S. N., Kohl, H. W., III, Barlow, C. E., Paffenbarger, R. S., Jr., Gibbons, L. W., & Macera, C. A. (1995). Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *Journal of the American Medical Association, 273,* 1093–1098. http://dx.doi.org/10.1001/jama.1995.03520380029031

Bos, M. W., Dijksterhuis, A., & van Baaren, R. B. (2008). On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology, 44,* 1114–1120. http://dx.doi.org/10.1016/j.jesp.2008.01.001

Boureau, Y. L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology, 36,* 74–97. http://dx.doi.org/10.1038/npp.2010.151

Boyer, P. (2006). Why ritualized behavior? Precaution systems and action-parsing in developmental, pathological and cultural rituals. *Behavioral and Brain Sciences, 29,* 1–56.

Broman, C. L. (1995). Leisure-time physical activity in an African-American population. *Journal of Behavioral Medicine, 18,* 341–353. http://dx.doi.org/10.1007/BF01857659

Carruthers, P. (2009). An architecture for dual reasoning. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 1–35). Oxford, England: New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199230167.003.0005

Cartoni, E., Puglisi-Allegra, S., & Baldassarre, G. (2013). The three principles of action: A Pavlovian-instrumental transfer hypothesis. *Frontiers in Behavioral Neuroscience, 7,* 153. http://dx.doi.org/10.3389/fnbeh.2013.00153

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality—social, clinical, and health psychology. *Psychological Bulletin, 92,* 111–135. http://dx.doi.org/10.1037/0033-2909.92.1.111

Cecchini, M., Sassi, F., Lauer, J. A., Lee, Y. Y., Guajardo-Barron, V., & Chisholm, D. (2010). Tackling of unhealthy diets, physical inactivity, and obesity: Health effects and cost-effectiveness. *The Lancet, 376,* 1775–1784. http://dx.doi.org/10.1016/S0140-6736(10)61514-0

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York, NY: Guilford Press.

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.

Chaloupka, F. J., & Grossman, M. (1996). *Price, tobacco control policies and youth smoking*. National Bureau of Economic Research Working Paper No. W5740.

Chaloupka, F. J., Grossman, M., Bickel, W. K., & Saffer, H. (1999). Introduction. In F. J. Chaloupka, M. Grossman, W. K. Bickel, & H. Saffer (Eds.), *The economic analysis of substance use and abuse: An integration of economic and behavioral economic research* (pp. 1–12). Chicago, IL: University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226100494.001.0001

Chaloupka, F. J., Grossman, M., & Saffer, H. (2002). The effects of price on alcohol consumption and alcohol-related problems. *Alcohol Research & Health, 26,* 22–34.

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition, 113,* 350–364. http://dx.doi.org/10.1016/j.cognition.2008.06.014

Choi, J. J., Laibson, D., Madrian, B., & Metrick, A. (2002). Defined contribution pensions: Plan rules, participant decisions, and the path of least resistance. In J. M. Poterba (Ed.), *Tax policy and the economy* (pp. 67–113). Cambridge, MA: MIT Press.

Choi, J. J., Laibson, D., Madrian, B., & Metrick, A. (2003). Optimal defaults. *The American Economic Review, 93,* 180–185. http://dx.doi.org/10.1257/000282803321947010

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science, 12,* 105–109. http://dx.doi.org/10.1111/1467-8721.01242

Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. New York, NY: Harper Business.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55,* 591–621. http://dx.doi.org/10.1146/annurev.psych.55.090902.142015

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct. *Advances in Experimental Social Psychology, 24,* 201–234. http://dx.doi.org/10.1016/S0065-2601(08)60330-5

Cioffi, D., & Garner, R. (1996). On doing the decision: Effects of active versus passive commitment and self-perception. *Personality and Social Psychology Bulletin, 22,* 133–147. http://dx.doi.org/10.1177/0146167296222003

Clapp, J. D., & McDonnell, A. (2000). The relationship of perceptions of alcohol promotion and peer drinking norms to alcohol problems reported by college students. *Journal of College Student Development, 41,* 19–26.

Clark, D. M., & Fairburn, C. G. (Eds.). (1997). *Science and practice of cognitive behaviour therapy*. Oxford, United Kingdom: Oxford University Press.

Cohen, J. B., Pham, M. T., & Andrade, E. B. (2008). The nature and role of affect in consumer behavior. In C. P. Haugtvedt, P. Herr, & F. Kardes (Eds.), *Handbook of consumer psychology* (pp. 297–348). New York, NY: Erlbaum.

Colman, G., Grossman, M., & Joyce, T. (2003). The effect of cigarette excise taxes on smoking before, during and after pregnancy. *Journal of Health Economics, 22,* 1053–1072. http://dx.doi.org/10.1016/j.jhealeco.2003.06.003

Critchley, H. D. (2005). Neural mechanisms of autonomic, affective, and cognitive integration. *The Journal of Comparative Neurology, 493,* 154–166. http://dx.doi.org/10.1002/cne.20749

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7,* 189–195. http://dx.doi.org/10.1038/nn1176

Curtis, V. A., Danquah, L. O., & Aunger, R. V. (2009). Planned, motivated and habitual hygiene behaviour: An eleven country review. *Health Education Research, 24,* 655–673. http://dx.doi.org/10.1093/her/cyp002

Curtis, V. A., Garbrah-Aidoo, N., & Scott, B. (2007). Ethics in public health research: Masters of marketing: Bringing private sector skills to public health partnerships. *American Journal of Public Health, 97,* 634–641. http://dx.doi.org/10.2105/AJPH.2006.090589

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: GP Putnam.

Damasio, A. R. (2000). *The feeling of what happens*. New York, NY: Harcourt Brace & Company.

Darnton, A. (2008). *Reference Report: An overview of behaviour change models and their uses*. London: Government Social Research Unit. Retrieved from http://www.gsr.gov.uk/resources/behaviour_change_review.asp

Davidson, K. W., Goldstein, M., Kaplan, R. M., Kaufmann, P. G., Knatterud, G. L., Orleans, C. T., . . . Whitlock, E. P. (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine, 26,* 161–171. http://dx.doi.org/10.1207/S15324796ABM2603_01

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711. http://dx.doi.org/10.1038/nn1560

Dayan, P. (2008). The role of value systems in decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Decision making, the human mind, and implications for institutions* (pp. 51–70). Frankfurt, Germany: MIT Press.

Dayan, P. (2009). Goal-directed control and its antipodes. *Neural Networks, 22,* 213–219. http://dx.doi.org/10.1016/j.neunet.2009.03.004

Dayan, P. (2012). Models of value and choice. In T. Sharot & R. Dolan (Eds.), *Neuroscience of preference and choice* (pp. 33–52). Waltham: Academic Press. http://dx.doi.org/10.1016/B978-0-12-381431-9.00002-4

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

Dayan, P., & Niv, Y. (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology, 18,* 185–196. http://dx.doi.org/10.1016/j.conb.2008.08.003

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Networks, 19,* 1153–1160. http://dx.doi.org/10.1016/j.neunet.2006.03.002

Dayan, P., & Seymour, B. (2009). Values and actions in aversion. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 175–191). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-374176-9.00012-9

Deci, E., & Ryan, R. (Eds.). (2002). *Handbook of self-determination research*. Rochester, NY: University of Rochester Press.

Delgado, M. R., Gillis, M. M., & Phelps, E. A. (2008). Regulating the expectation of reward via cognitive strategies. *Nature Neuroscience, 11,* 880–881. http://dx.doi.org/10.1038/nn.2141

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313,* 684–687. http://dx.doi.org/10.1126/science.1128356

Department of Health. (2004). *Choosing health: Making healthy choices easier*. Public Health White Paper (The Stationery Office, Department of Health).

de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., & Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *The Journal of Neuroscience, 32,* 12066–12075. http://dx.doi.org/10.1523/JNEUROSCI.1088-12.2012

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. In R. Gallistel (Ed.), *Stevens' handbook of experimental psychology* (Vol. 3, pp. 497–534). New York, NY: Wiley. http://dx.doi.org/10.1002/0471214426.pas0312

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 33, pp. 1–40). San Diego, CA: Academic.

Dijksterhuis, A., Bargh, J. A., & Miedema, J. (2000). Of men and mackerels: Attention, subjective experience, and automatic social behavior. In H. Bless & J. P. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior* (pp. 37–51). Philadelphia, PA: Psychology Press.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science, 311,* 1005–1007. http://dx.doi.org/10.1126/science.1121629

Dijksterhuis, A., Chartrand, T. L., & Aarts, H. (2007). Effects of priming and perception on social behavior and goal pursuit. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 51–132). Philadelphia, PA: Psychology Press.

Dolan, P., Elliott, A., Metcalfe, R., & Vlaev, I. (2012a). Influencing financial behavior: From changing minds to changing contexts. *Journal of Behavioral Finance, 13,* 126–142. http://dx.doi.org/10.1080/15427560.2012.680995

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012b). Influencing behaviour: The mindspace way. *Journal of Economic Psychology, 33,* 264–277. http://dx.doi.org/10.1016/j.joep.2011.10.009

Dolan, P., Hallsworth, M., Halpern, D., King, D., & Vlaev, I. (2010). *MINDSPACE: Influencing behaviour through public policy*. Report for the United Kingdom Cabinet Office. Institute for Government.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80,* 312–325. http://dx.doi.org/10.1016/j.neuron.2013.09.007

Duflo, E., Kremer, M., & Robinson, J. (2010). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *The American Economic Review, 101,* 2350–2390.

Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal Applied Economics, 3,* 1–34. http://dx.doi.org/10.1257/app.3.1.1

Durantini, M. R., Albarracín, D., Mitchell, A. L., Earl, A. N., & Gillette, J. C. (2006). Conceptualizing the influence of social agents of behavior change: A meta-analysis of the effectiveness of HIV-prevention interventionists for different groups. *Psychological Bulletin, 132,* 212–248.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54,* 5–12. http://dx.doi.org/10.1037/0022-3514.54.1.5

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931. http://dx.doi.org/10.1037/0033-295X.112.4.912

Estes, W. K., & Skinner, B. F. (1941). Some quantitative properties of anxiety. *Journal of Experimental Psychology, 29,* 390–400. http://dx.doi.org/10.1037/h0062283

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278. http://dx.doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8,* 223–241. http://dx.doi.org/10.1177/1745691612460685

Everitt, B. J., Dickinson, A., & Robbins, T. W. (2001). The neuropsychological basis of addictive behaviour. *Brain Research Reviews, 36,* 129–138. http://dx.doi.org/10.1016/S0165-0173(01)00088-1

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison Wesley.

Fisher, J. D., Fisher, W. A., Bryan, A. D., & Misovich, S. J. (2002). Information-motivation-behavioral skills model-based HIV risk behavior change intervention for inner-city high school youth. *Health Psychology, 21,* 177–186. http://dx.doi.org/10.1037/0278-6133.21.2.177

Fiske, S. T. (2010). *Social beings: Core motives in social psychology*. New York, NY: Wiley.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731. http://dx.doi.org/10.1037/0033-2909.132.5.692

Gentile, E., & Imberman, S. A. (2009). *Dressed for success: Do school uniforms improve student behavior, attendance, and achievement?* (Working Paper 09-03). Department of Economics, University of Houston.

Gerrard, M., Gibbons, F. X., Lane, D. J., & Stock, M. L. (2005). Social comparison and smoking cessation: Comparison level predicts success for adult smokers. *Health Psychology, 24,* 623–629. http://dx.doi.org/10.1037/0278-6133.24.6.623

Gerrard, M., Gibbons, F. X., Reis-Bergan, M., Trudeau, L., Vande Lune, L., & Buunk, B. P. (2002). Inhibitory effects of drinker and nondrinker prototypes on adolescent alcohol consumption. *Health Psychology, 21,* 601–609. http://dx.doi.org/10.1037/0278-6133.21.6.601

Gibbons, F. X., & Gerrard, M. (1995). Predicting young adults' health risk behavior. *Journal of Personality and Social Psychology, 69,* 505–517. http://dx.doi.org/10.1037/0022-3514.69.3.505

Gibbons, F. X., Gerrard, M., Blanton, H., & Russell, D. W. (1998). Reasoned action and social reaction: Willingness and intention as independent predictors of health risk. *Journal of Personality and Social Psychology, 74,* 1164–1180. http://dx.doi.org/10.1037/0022-3514.74.5.1164

Gibbons, F. X., Gerrard, M., Lane, D. J., Mahler, H. I. M., & Kulik, J. A. (2005). Using UV photography to reduce use of tanning booths: A test of cognitive mediation. *Health Psychology, 24,* 358–363. http://dx.doi.org/10.1037/0278-6133.24.4.358

Gibbons, F. X., Lane, D. J., Gerrard, M., Pomery, E. A., & Lautrup, C. L. (2002). Drinking and driving: A prospective assessment of the relation between risk cognitions and risk behavior. *Risk, Decision and Policy, 7,* 267–283. http://dx.doi.org/10.1017/S1357530902000601

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. http://dx.doi.org/10.1037/0033-295X.103.4.650

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review, 115,* 230–239. http://dx.doi.org/10.1037/0033-295X.115.1.230

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595. http://dx.doi.org/10.1016/j.neuron.2010.04.016

Glasgow, R. E., Klesges, L. M., Dzewaltowski, D. A., Bull, S. S., & Estabrooks, P. (2004). The future of health behavior change research: What is needed to improve translation of research into health promotion practice? *Annals of Behavioral Medicine, 27,* 3–12. http://dx.doi.org/10.1207/s15324796abm2701_2

Glimcher, P. W., Camerer, C. F., Fehr, E., & Poldrack, R. A. (Eds.). (2009). *Neuroeconomics: Decision making and the brain*. New York, NY: Academic Press.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research, 35,* 472–482. http://dx.doi.org/10.1086/586910

Gollwitzer, P. M. (1993). Goal achievement: The role of intentions. *European Review of Social Psychology, 4,* 141–185. http://dx.doi.org/10.1080/14792779343000059

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503. http://dx.doi.org/10.1037/0003-066X.54.7.493

Gollwitzer, P. M., & Brandstatter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology, 73,* 186–199. http://dx.doi.org/10.1037/0022-3514.73.1.186

Gottlieb, J., & Balan, P. (2010). Attention as a decision in information space. *Trends in Cognitive Sciences, 14,* 240–248. http://dx.doi.org/10.1016/j.tics.2010.03.001

Gray, J. B. (2008). Framing: A communication strategy for the medical encounter. *Journal of Community Health, 1,* 422–430. http://dx.doi.org/10.1179/cih.2008.1.4.422

Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 4115–4120. http://dx.doi.org/10.1073/pnas.062381899

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior, 26,* 245–256. http://dx.doi.org/10.1016/j.evolhumbehav.2005.01.002

Halpern, S. D., Ubel, P. A., & Asch, D. A. (2007). Harnessing the power of default options to improve health care. *The New England Journal of Medicine, 357,* 1340–1344. http://dx.doi.org/10.1056/NEJMsb071595

Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health Psychology, 28,* 404–413. http://dx.doi.org/10.1037/a0014399

Hausenblas, H. A., Caron, A. V., & Mack, D. E. (1997). Application of the theories of reasoned action and planned behavior to exercise behavior: A meta-analysis. *Journal of Sport & Exercise Psychology, 19,* 36–51.

Hawkes, N. (2004). Give up or you'll clog up, says charity. *The Times,* January 1.

Heaney, C. A., & Israel, B. A. (1997). Social networks and social support. In K. Glanz, F. M. Lewis, & B. K. Rimer (Eds.), *Health behavior and health education: Theory, research, and practice* (pp. 179–205). San Francisco, CA: Jossey-Bass.

Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion, 11,* 101–120. http://dx.doi.org/10.1007/BF00992338

Heller, R. F., & Page, J. (2002). A population perspective to evidence based medicine: "evidence for population health". *Journal of Epidemiology and Community Health, 56,* 45–47. http://dx.doi.org/10.1136/jech.56.1.45

Hershberger, W. A. (1986). An approach through the looking-glass. *Animal Learning & Behavior, 14,* 443–451. http://dx.doi.org/10.3758/BF03200092

Higgins, S. T., Heil, S. H., & Lussier, J. P. (2004). Clinical implications of reinforcement as a determinant of substance use disorders. *Annual Review of Psychology, 55,* 431–461. http://dx.doi.org/10.1146/annurev.psych.55.090902.142033

Hirsh, R. (1974). The hippocampus and contextual retrieval of information from memory: A theory. *Behavioral Biology, 12,* 421–444. http://dx.doi.org/10.1016/S0091-6773(74)92231-7

Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit. Nonconscious effects of scent on cognition and behavior. *Psychological Science, 16,* 689–693. http://dx.doi.org/10.1111/j.1467-9280.2005.01597.x

Houser, D., Reiley, D., & Urbancic, M. (2008). *Checking out temptation: A natural experiment with purchases at the grocery register* (Working Paper). George Mason University.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264. http://dx.doi.org/10.1037/0033-295X.110.2.220

Johnson, E. J., Bellman, S., & Lohse, G. L. (2002). Defaults, framing and privacy: Why opting in-opting out. *Marketing Letters, 13,* 5–15. http://dx.doi.org/10.1023/A:1015044207315

Johnson, E. J., & Goldstein, D. (2003). Medicine. Do defaults save lives? *Science, 302,* 1338–1339. http://dx.doi.org/10.1126/science.1091721

Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty, 7,* 35–51. http://dx.doi.org/10.1007/BF01065313

Jones, J. L., & Leary, M. R. (1994). Effects of appearance-based admonitions against sun exposure on tanning intentions in young adults. *Health Psychology, 13,* 86–90. http://dx.doi.org/10.1037/0278-6133.13.1.86

Judah, G., Aunger, R., Schmidt, W. P., Michie, S., Granger, S., & Curtis, V. (2009). Experimental pretesting of hand-washing interventions in a natural setting. *American Journal of Public Health, 99*(Suppl. 2), S405–S411. http://dx.doi.org/10.2105/AJPH.2009.164160

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697–720. http://dx.doi.org/10.1037/0003-066X.58.9.697

Kahneman, D., & Frederick, S. (2007). Frames and brains: Elicitation and control of response tendencies. *Trends in Cognitive Sciences, 11,* 45–46. http://dx.doi.org/10.1016/j.tics.2006.11.007

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values and frames*. New York, NY: Cambridge University Press and the Russell Sage Foundation.

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. *Science, 322,* 1681–1685. http://dx.doi.org/10.1126/science.1161405

Kinmonth, A. L., Wareham, N. J., Hardeman, W., Sutton, S., Prevost, A. T., Fanshawe, T., . . . Griffin, S. J. (2008). Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk group in primary care (ProActive UK): A randomised trial. *The Lancet, 371,* 41–48. http://dx.doi.org/10.1016/S0140-6736(08)60070-7

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61,* 140–151. http://dx.doi.org/10.1016/j.neuron.2008.11.027

Kreps, D. M. (1990). *A course in microeconomic theory*. New York, NY: Harvester Wheatsheaf.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112,* 443–478. http://dx.doi.org/10.1162/003355397555253

Lally, P., Chipperfield, A., & Wardle, J. (2008). Healthy habits: Efficacy of simple advice on weight control based on a habit-formation model. *International Journal of Obesity, 32,* 700–707. http://dx.doi.org/10.1038/sj.ijo.0803771

Lally, P., van Jaarsveld, C., Potts, H., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40,* 998–1009. http://dx.doi.org/10.1002/ejsp.674

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155–184. http://dx.doi.org/10.1146/annurev.neuro.23.1.155

Levav, J., & Fitzsimons, G. J. (2006). When questions change behavior: The role of ease of representation. *Psychological Science, 17,* 207–213. http://dx.doi.org/10.1111/j.1467-9280.2006.01687.x

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

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review, 90,* 426–432. http://dx.doi.org/10.1257/aer.90.2.426

Mackintosh, N. J. (1983). *Conditioning and associative learning*. New York, NY: Oxford University Press.

Madrian, B. C., & Shea, D. F. (2001). The power of suggestion: Inertia in 401(k) participation and savings behavior. *The Quarterly Journal of Economics, 116,* 1149–1187. http://dx.doi.org/10.1162/003355301753265543

Mahler, H. I. M., Kulik, J. A., Gibbons, F. X., Gerrard, M., & Harrell, J. (2003). Effects of appearance-based interventions on sun protection intentions and self-reported behaviors. *Health Psychology, 22,* 199–209. http://dx.doi.org/10.1037/0278-6133.22.2.199

Mann, T., & Ward, A. (2007). Attention, self-control, and health behaviors. *Current Directions in Psychological Science, 16,* 280–283. http://dx.doi.org/10.1111/j.1467-8721.2007.00520.x

Manning, W. G., Blumberg, L., & Moulton, L. H. (1995). The demand for alcohol: The differential response to price. *Journal of Health Economics, 14,* 123–148. http://dx.doi.org/10.1016/0167-6296(94)00042-3

Marteau, T. M., Ashcroft, R. E., & Oliver, A. (2009). Using financial incentives to achieve healthy behaviour. *British Medical Journal, 338,* b1415. http://dx.doi.org/10.1136/bmj.b1415

Marteau, T. M., & Lerman, C. (2001). Genetic risk and behavioural change. *British Medical Journal, 322,* 1056–1059. http://dx.doi.org/10.1136/bmj.322.7293.1056

Marteau, T. M., & Weinman, J. A. (2006). Self-regulation and the behavioural response to DNA risk information: A theoretical analysis and framework for future research. *Social Science & Medicine, 62,* 1360–1368. http://dx.doi.org/10.1016/j.socscimed.2005.08.005

Martin, S. J., Bassi, S., & Dunbar-Rees, R. (2012). Commitments, norms and custard creams: A social influence approach to reducing did not attends (DNAs). *Journal of the Royal Society of Medicine, 105,* 101–104. http://dx.doi.org/10.1258/jrsm.2011.110250

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *The Journal of Neuroscience, 27,* 5796–5804. http://dx.doi.org/10.1523/JNEUROSCI.4246-06.2007

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science, 306,* 503–507. http://dx.doi.org/10.1126/science.1100907

McConnell, A. R., & Rydell, R. J. (2014). The Systems of Evaluation Model: A dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 204–217). New York, NY: Guilford Press.

McCormick, M., & McElroy, T. (2009). Healthy choices in context: How contextual cues can influence the persuasiveness of framed health messages. *Judgment and Decision Making, 4,* 248–255.

Melvin, K. B., & Anson, J. E. (1969). Facilitative effects of punishment on aggressive behavior in Siamese fighting fish. *Psychonomic Science, 14,* 89–90. http://dx.doi.org/10.3758/BF03332718

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Pt. 1. Basic mechanisms. *Psychological Review, 104,* 3–65. http://dx.doi.org/10.1037/0033-295X.104.1.3

Michie, S., & Abraham, C. (2004). Interventions to change health behaviours: evidence-based or evidence-inspired? *Psychology & Health, 19,* 29–49. http://dx.doi.org/10.1080/0887044031000141199

Michie, S., Ashford, S., Sniehotta, F. F., Dombrowski, S. U., Bishop, A., & French, D. P. (2011). A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: The CALO-RE taxonomy. *Psychology & Health, 26,* 1479–1498. http://dx.doi.org/10.1080/08870446.2010.540664

Michie, S., Johnston, M., Francis, J., Hardeman, W., & Eccles, M. (2008). From theory to intervention: Mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology, 57,* 660–680. http://dx.doi.org/10.1111/j.1464-0597.2008.00341.x

Michie, S., van Stralen, M. M., & West, R. (2011). The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science: IS, 6,* 42. http://dx.doi.org/10.1186/1748-5908-6-42

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202. http://dx.doi.org/10.1146/annurev.neuro.24.1.167

Morse, W. H., Mead, R. N., & Kelleher, R. T. (1967). Modulation of elicited behavior by a fixed-interval schedule of electric shock presentation. *Science, 157,* 215–217. http://dx.doi.org/10.1126/science.157.3785.215

Moser, K., Patnick, J., & Beral, V. (2009). Inequalities in reported use of breast and cervical screening in Great Britain: Analysis of cross sectional survey data. *British Medical Journal, 338,* b2025. http://dx.doi.org/10.1136/bmj.b2025

Moss, A. C., & Albery, I. P. (2009). A dual-process model of the alcohol-behavior link for social drinking. *Psychological Bulletin, 135,* 516–530. http://dx.doi.org/10.1037/a0015991

Moss, A. C., & Albery, I. P. (2010). Are alcohol expectancies associations, propositions, or elephants? A reply to Wiers and Stacy (2010). *Psychological Bulletin, 136,* 17–20. http://dx.doi.org/10.1037/a0018087

Mukherjee, K. (2010). A dual system model of preferences under risk. *Psychological Review, 117,* 243–255. http://dx.doi.org/10.1037/a0017884

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472–489. http://dx.doi.org/10.1037/0033-295X.110.3.472

National Social Marketing Foundation. (2006). *National Health-related Campaigns Review. Report 7*. London: National Social Marketing Foundation.

Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits: A repeated performance. *Current Directions in Psychological Science, 15,* 198–202. http://dx.doi.org/10.1111/j.1467-8721.2006.00435.x

Niv, Y., Joel, D., & Dayan, P. (2006). A normative perspective on motivation. *Trends in Cognitive Sciences, 10,* 375–381. http://dx.doi.org/10.1016/j.tics.2006.06.010

Norman, P., Abraham, C., & Conner, M. (Eds.). (2000). *Understanding and changing health behaviour: From health beliefs to self-regulation* (2nd ed.). Amsterdam: Harwood Academic.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9,* 242–249. http://dx.doi.org/10.1016/j.tics.2005.03.010

O'Donoghue, T., & Rabin, M. (1999). Doing it now or doing it later. *The American Economic Review, 89,* 103–124. http://dx.doi.org/10.1257/aer.89.1.103

O'Keefe, D. J., & Jensen, J. D. (2007). The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: A meta-analytic review. *Journal of Health Communication, 12,* 623–644. http://dx.doi.org/10.1080/10810730701615198

O'Keefe, D. J., & Jensen, J. D. (2008). Do loss-framed persuasive messages engender greater message processing than do gain-framed messages? A meta-analytic review. *Communication Studies, 59,* 51–67. http://dx.doi.org/10.1080/10510970701849388

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Oxford University Press.

Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology, 29,* 374–383. http://dx.doi.org/10.1037/a0019596

Ouellette, J. A., Hessling, R., Gibbons, F. X., Reis-Bergan, M., & Gerrard, M. (2005). Using images to increase exercise behavior: Prototypes versus possible selves. *Personality and Social Psychology Bulletin, 31,* 610–620. http://dx.doi.org/10.1177/0146167204271589

Ouellette, J. A., & Wood, W. (1998). Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological Bulletin, 124,* 54–74. http://dx.doi.org/10.1037/0033-2909.124.1.54

Park, C. W., Jun, S. Y., & MacInnis, D. J. (2000). Choosing what I want versus rejecting what I do not want: An application of decision framing to product option choice decisions. *Journal of Marketing Research, 37,* 187–202. http://dx.doi.org/10.1509/jmkr.37.2.187.18731

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., . . . Wilmore, J. H. (1995a). Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of the American Medical Association, 273,* 402–407. http://dx.doi.org/10.1001/jama.1995.03520290054029

Paul-Ebhohimhen, V., & Avenell, A. (2008). Systematic review of the use of financial incentives in treatments for obesity and overweight. *Obesity Reviews, 9,* 355–367. http://dx.doi.org/10.1111/j.1467-789X.2007.00409.x

Pearce, D. W. (Ed.). (1986). *The MIT dictionary of modern economics* (3rd ed.). Cambridge, MA: MIT Press.

Perkins, H. W. (2004). *The social norms approach to prevention*. San Francisco, CA: Jossey-Bass.

Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming. *International Journal of the Addictions, 21,* 961–976.

Perkins, H. W., & Craig, D. W. (2006). A successful social norms campaign to reduce alcohol misuse among college student-athletes. *Journal of Studies on Alcohol, 67,* 880–889.

Petty, R. E., & Briñol, P. (2014). The elaboration likelihood and meta-cognitive models of attitudes: Implications for prejudice, the self, and beyond. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 172–187). New York, NY: Guilford Press.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4612-4964-1

Phelps, E. A. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology, 57,* 27–53. http://dx.doi.org/10.1146/annurev.psych.56.091103.070234

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., . . . David, A. S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389,* 495–498. http://dx.doi.org/10.1038/39051

Prochaska, J. O., & DiClemente, C. C. (1984). *The transtheoretical approach: Crossing the traditional boundaries of change*. Homewood, IL: Irwin.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9,* 545–556. http://dx.doi.org/10.1038/nrn2357

Rivis, A. J., & Sheeran, P. (2004). Descriptive norms as an additional predictor in the theory of planned behaviour: A meta-analysis. *Current Psychology, 22,* 264–280.

Rogers, R. W. (1983). Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation. In J. T. Cacioppo & R. E. Petty (Eds.), *Social psychophysiology: A sourcebook* (pp. 494–507). New York, NY: Guilford Press.

Roll, J. M., Higgins, S. T., & Badger, G. J. (1996). An experimental comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. *Journal of Applied Behavior Analysis, 29,* 495–505. http://dx.doi.org/10.1901/jaba.1996.29-495

Rolls, E. T. (2005). *Emotion explained*. Oxford: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780198570035.001.0001

Rolls, E. T. (2014). *Emotion and decision-making explained*. Oxford: Oxford University Press.

Rosenstock, I. (1974). Historical origins of the health belief model. *Health Education Behavior, 2,* 328–335. http://dx.doi.org/10.1177/109019817400200403

Rothman, A. J., Hertel, A. W., Baldwin, A. S., & Bartels, R. (2007). Understanding the determinants of health behavior change: Integrating theory and practice. In J. Shah & W. Gardner (Eds.), *Handbook of motivation science* (pp. 494–507). New York, NY: Guilford Press.

Ruiter, R. A. C., Kok, G., Verplanken, B., & Brug, J. (2001). Evoked fear and effects of appeals on attitudes to performing breast self-examination: An information-processing perspective. *Health Education Research, 16,* 307–319. http://dx.doi.org/10.1093/her/16.3.307

Rychetnik, L., Frommer, M., Hawe, P., & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health, 56,* 119–127. http://dx.doi.org/10.1136/jech.56.2.119

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7–59. http://dx.doi.org/10.1007/BF00055564

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science, 18,* 429–434. http://dx.doi.org/10.1111/j.1467-9280.2007.01917.x

Schwarzer, R. (1999). Self-regulatory processes in the adoption and maintenance of health behaviors. *Journal of Health Psychology, 4,* 115–127. http://dx.doi.org/10.1177/135910539900400208

Schweitzer, M., Hershey, J. C., & Asch, D. A. (1996). Individual choice in spending accounts. Can we rely on employees to choose well? *Medical Care, 34,* 583–593. http://dx.doi.org/10.1097/00005650-199606000-00008

Seymour, B., & Dolan, R. (2008). Emotion, decision making, and the amygdala. *Neuron, 58,* 662–671. http://dx.doi.org/10.1016/j.neuron.2008.05.020

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience, 8,* 300–311. http://dx.doi.org/10.1038/nrn2119

Shah, J. Y., & Gardner, W. L. (2008). *Handbook of motivational science*. New York, NY: Guilford Press.

Sheeran, P. (2002). Intention-behaviour relations: A conceptual and empirical review. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 1–36). London: Wiley. http://dx.doi.org/10.1080/14792772143000003

Sheeran, P., Abraham, C., & Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: A meta-analysis. *Psychological Bulletin, 125,* 90–132. http://dx.doi.org/10.1037/0033-2909.125.1.90

Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin, 31,* 87–98. http://dx.doi.org/10.1177/0146167204271308

Shumaker, S. A., Schron, E., Ockene, J., & McBee, W. L. (Eds.). (2008). *The handbook of health behavior change* (2nd ed.). New York, NY: Springer.

Sieverding, M., Decker, S., & Zimmermann, F. (2010). Information about low participation in cancer screening demotivates other people. *Psychological Science, 21,* 941–943. http://dx.doi.org/10.1177/0956797610373936

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138. http://dx.doi.org/10.1037/h0042769

Simon, H. A. (1992). *Economics, bounded rationality and the cognitive revolution*. Aldershot, England: Elgar.

Sindelar, J. L. (2008). Paying for performance: The power of incentives over habits. *Health Economics, 17,* 449–451. http://dx.doi.org/10.1002/hec.1350

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303,* 1157–1162. http://dx.doi.org/10.1126/science.1093535

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. http://dx.doi.org/10.1037/0033-2909.119.1.3

Slovic, P., Finucane, M., Peters, E., & McGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgement* (pp. 397–420). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511808098.025

Sniehotta, F. F. (2009). An experimental test of the theory of planned behaviour. *Applied Psychology, 1,* 257–270.

Sobal, J., & Wansink, B. (2007). Kitchenscapes, tablescapes, platescapes, and foodscapes: Influence of microscale built environments on food intake. *Environment and Behavior, 39,* 124–142. http://dx.doi.org/10.1177/0013916506295574

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220–247. http://dx.doi.org/10.1207/s15327957pspr0803_1

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge: New York, NY: Cambridge University Press.

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human pavlovian-instrumental transfer. *The Journal of Neuroscience, 28,* 360–368. http://dx.doi.org/10.1523/JNEUROSCI.4028-07.2008

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee savings. *Journal of Political Economy, 112,* S164–S187. http://dx.doi.org/10.1086/380085

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thornton, R. L. (2008). The demand for, and impact of, learning HIV status. *The American Economic Review, 98,* 1829–1863. http://dx.doi.org/10.1257/aer.98.5.1829

Triandis, H. C. (1977). *Interpersonal behavior*. Monterey, CA: Brooks/Cole.

Trope, Y., & Fishbach, A. (2000). Counteractive self-control in overcoming temptation. *Journal of Personality and Social Psychology, 79,* 493–506.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79,* 281–299. http://dx.doi.org/10.1037/h0032955

Uitenbroek, D. G., Kerekovska, A., & Festchieva, N. (1996). Health lifestyle behaviour and socio-demographic characteristics. A study of Varna, Glasgow and Edinburgh. *Social Science & Medicine, 43,* 367–377. http://dx.doi.org/10.1016/0277-9536(95)00399-1

Verplanken, B., Friborg, O., Wang, C. E., Trafimow, D., & Woolf, K. (2007). Mental habits: Metacognitive reflection on negative self-thinking. *Journal of Personality and Social Psychology, 92,* 526–541. http://dx.doi.org/10.1037/0022-3514.92.3.526

Verplanken, B., & Velsvik, R. (2008). Habitual negative body image thinking as psychological risk factor in adolescents. *Body Image, 5,* 133–140. http://dx.doi.org/10.1016/j.bodyim.2007.11.001

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. A. (2011). Does the brain calculate value? *Trends in Cognitive Sciences, 15,* 546–554. http://dx.doi.org/10.1016/j.tics.2011.09.008

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: A randomized trial. *Journal of the American Medical Association, 300,* 2631–2637. http://dx.doi.org/10.1001/jama.2008.804

Wadden, T. A., Volger, S., Sarwer, D. B., Vetter, M. L., Tsai, A. G., Berkowitz, R. I., . . . Moore, R. H. (2011). A two-year randomized trial of obesity treatment in primary care practice. *The New England Journal of Medicine, 365,* 1969–1979. http://dx.doi.org/10.1056/NEJMoa1109220

Wakefield, M. A., Loken, B., & Hornik, R. C. (2010). Use of mass media campaigns to change health behaviour. *The Lancet, 376,* 1261–1271. http://dx.doi.org/10.1016/S0140-6736(10)60809-4

Wansink, B., & Cheney, M. M. (2005). Super Bowls: Serving bowl size and food consumption. *Journal of the American Medical Association, 293,* 1727–1728.

Wansink, B., & van Ittersum, K. (2011, in press). The perils of plate size: Waist, waste, and wallet. *Journal of Marketing*.

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132,* 249–268. http://dx.doi.org/10.1037/0033-2909.132.2.249

Werner, C. M., Turner, J., Shipman, K., Twitchell, S. F., Dickson, B. R., Bruschke, G. V., & von Bismarck, W. B. (1995). Commitment, behavior, and attitude change: An analysis of voluntary recycling. *Journal of Environmental Psychology, 15,* 197–208. http://dx.doi.org/10.1016/0272-4944(95)90003-9

West, R. (2006). *Theory of addiction*. Oxford: Wiley-Blackwell.

Whitby, M., McLaws, M.-L., & Ross, M. W. (2006). Why healthcare workers don't wash their hands: A behavioral explanation. *Infection Control and Hospital Epidemiology, 27,* 484–492. http://dx.doi.org/10.1086/503335

White, N. M., Packard, M. G., & McDonald, R. J. (2013). Dissociation of memory systems: The story unfolds. *Behavioral Neuroscience, 127,* 813–834. http://dx.doi.org/10.1037/a0034859

Williams, B. R., Bezner, J., Chesbro, S. B., & Leavitt, R. (2006). The effect of a walking program on perceived benefits and barriers to exercise in postmenopausal African American women. *Journal of Geriatric Physical Therapy, 29,* 43–49. http://dx.doi.org/10.1519/00139143-200608000-00001

Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. *Journal of the Experimental Analysis of Behavior, 12,* 511–520. http://dx.doi.org/10.1901/jeab.1969.12-511

Wolpe, J. (1990). *The practice of behavior therapy* (4th ed.). New York, NY: Pergamon Press.

Wood, W., & Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review, 114,* 843–863. http://dx.doi.org/10.1037/0033-295X.114.4.843

Wood, W., & Neal, D. T. (2009). The habitual consumer. *Journal of Consumer Psychology, 19,* 579–592. http://dx.doi.org/10.1016/j.jcps.2009.08.003

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology, 83,* 1281–1297. http://dx.doi.org/10.1037/0022-3514.83.6.1281

Wood, W., Tam, L., & Witt, M. G. (2005). Changing circumstances, disrupting habits. *Journal of Personality and Social Psychology, 88,* 918–933. http://dx.doi.org/10.1037/0022-3514.88.6.918

Worth, K., Sullivan, H., Hertel, A. W., Jeffery, R. W., & Rothman, A. J. (2005). Are there times when avoidance goals can be beneficial? A look at smoking cessation. *Basic and Applied Social Psychology, 27,* 107–116. http://dx.doi.org/10.1207/s15324834basp2702_2

Wryobeck, J., & Chen, Y. (2003). Using priming techniques to facilitate health behaviours. *Clinical Psychologist, 7,* 105–108. http://dx.doi.org/10.1080/13284200410001707553

Ybarra, O., & Trafimow, D. (1998). How priming the private self or the collective self affects the relative weights of attitudes and subjective norms. *Personality and Social Psychology Bulletin, 24,* 362–370. http://dx.doi.org/10.1177/0146167298244003

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–175. http://dx.doi.org/10.1037/0003-066X.35.2.151

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., & Berns, G. S. (2003). Human striatal response to salient nonrewarding stimuli. *The Journal of Neuroscience, 23,* 8092–8097.

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron, 42,* 509–517. http://dx.doi.org/10.1016/S0896-6273(04)00183-7

Received September 30, 2014
Revision received January 3, 2015
Accepted January 26, 2015 ■