This chapter was originally published in the book *Advances in Experimental Social Psychology, Vol. 52* published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who know you, and providing a copy to your institution’s administrator.

All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution’s website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

[http://www.elsevier.com/locate/permissionusematerial](http://www.elsevier.com/locate/permissionusematerial)


© Copyright 2015 Elsevier Inc.

Academic Press
CHAPTER TWO

A Biosocial Model of Affective Decision Making: Implications for Dissonance, Motivation, and Culture

Shinobu Kitayama1, Steven Tompson
University of Michigan, Ann Arbor, Michigan, USA
1Corresponding author: e-mail address: kitayama@umich.edu

Contents

1. Introduction 72
   1.1 Behavioral Conflict, Affect, and Motivation 72
   1.2 Dissonance Revolution 74
   1.3 Charting the Terrain 74
2. The Biosocial Model 77
   2.1 Key Propositions 77
   2.2 Moving Forward 81
3. Neural Substrates of Affective Decision Making 82
   3.1 Behavioral Conflict and Negative Arousal 82
   3.2 Making Decisions: Reward Processing and Response Selection 86
   3.3 Testing the Decision Mechanism 90
   3.4 Effort Justification 93
   3.5 Animal Dissonance 95
   3.6 Learning from Addiction 96
   3.7 Ego-Depletion 98
4. Top-Down Regulation 102
   4.1 Brain Mechanisms 102
   4.2 Action Orientation 104
   4.3 Arousal or Reduction? 106
5. Bottom-Up Regulation 108
   5.1 Brain Mechanisms 108
   5.2 Generalized Other and dACC Sensitivity: A Cultural Variation 111
   5.3 Culture, Face Priming, and Dissonance 115
   5.4 Self-Affirmation, Mortality Salience, and Other Bottom-Up Factors 120
6. Recursive Loop 123
   6.1 Aversive Consequences 123
   6.2 Misattribution 123
   6.3 Split Brain 124
Abstract

Drawing on recent advances in both neuroscience and animal behavior, we propose a biosocial model of affective decision making, which holds that when people face a conflict between two competing behavioral options (e.g., go vs. no-go, approach vs. avoidance), they develop a new affective disposition that resolves the conflict. This newly emerging affect will enable one to select a response while forming the basis for an elaborate cognition that justifies the selected response. The model reconceptualizes cognitive dissonance as fundamentally affective and involving both predecisional and postdecisional components. Furthermore, by postulating both top-down and bottom-up neural pathways to regulate the sensitivity to behavioral conflict, it integrates prior evidence on factors that moderate dissonance, including action orientation, self-affirmation, mortality salience, and culture. It also offers new insights into a disparate set of motivational phenomena including animal behaviors that mimic cognitive dissonance, sunk-cost fallacy, addiction, and ego-depletion. Lastly, the biosocial model has implications for how humans may be affectively and motivationally attached to symbols of culture. Directions for future research are discussed.

1. INTRODUCTION

1.1 Behavioral Conflict, Affect, and Motivation

People are conflicted in many different situations. Conflicts can happen when they face important decisions. Conflicts may also happen when they work on a demanding task such as climbing to reach the summit of a mountain. Similar conflicts will also be evident when one is confronting a loss of money, health, or a loved one through gambling, smoking, or using alcohol or illicit drugs. In all these cases, from decision making to mountain climbing, and from gambling to illicit drug use, a conflict arises because the two behavioral options are equally appealing for different reasons, but one course of action (choosing to work for one company; continuing to work on a demanding task or to gamble) precludes the other (choosing to work for the other company; stopping work on a demanding task or giving up gambling).

More often than not, the climbers continue to climb and the gamblers to gamble. Moreover, they often appear to be even more attracted to the
endeavor as a result of their desire to discontinue it. For example, the climbers appear to be more attracted than ever to reaching the summit as a result of experiencing the behavioral conflict. Could the climbers be more committed to climbing because they wanted to quit? More paradoxically, in some cases, people appear to be psychologically depleted (much like the climbers who are physically exhausted) after working on a highly demanding task, as if they have used up limited “muscle power” for self-control. Under such conditions, they find other impulses and temptations irresistible and difficult to control (Baumeister, Vohs, & Tice, 2007). Is this effect (called “ego-depletion”) related to the commitment of our climbers? Why are our climbers not “depleted” after having worked so hard to climb?

The thesis advanced in the current article is that the behavioral conflict associated with any demanding task or important decision is highly instrumental in determining subsequent decisions and behaviors, with an assortment of affective and motivational consequences. We propose that humans inherit from nonhuman animals certain brain circuitries that detect behavioral conflict. Once detected, the conflict initiates an active search for incentives associated with one of the response options to select and to pursue. Exactly what incentives are identified depends on what incentives are available and salient. The gamblers may be surrounded by a lot of attractive cues built into any Casino, and, likewise, the climbers may know how breathtaking the view from the summit would be. The gamblers may therefore become addicted to gambling, and the climbers may be more attracted to the endeavor. In contrast, people may be distracted if salient incentives are irrelevant to what they do. In such cases, they may appear being “depleted.” Although some of these phenomena, especially those involving post-decisional increases of commitment, have traditionally been studied under the rubric of cognitive dissonance (Aronson & Mills, 1959), we argue that the relevant brain circuitries responsible for these effects are largely sub-cortical and thus arguably precognitive. They can best be characterized as affective. We share these circuitries with rodents, birds, and nonhuman primates. The resulting affective dynamic is relevant in understanding the observations made above for the gamblers, the climbers, and the person who has been “depleted.” Moreover, this analysis will shed new light on the nature of all types of both decision making and decision rationalization.

Our thesis is couched initially in terms of cognitive dissonance theory. We will show that the above affective dynamic offers an important insight into a variety of phenomena covered in the dissonance literature of the last half century. Importantly, our analysis will go beyond the traditional
confines of dissonance theory. We will use the same theoretical model to understand several different phenomena, including addiction and self-control or the failure thereof. Importantly, the model is open to sociocultural influences and conditioning. This makes it possible to use the model to illuminate how people may be emotionally attached to symbolic systems of culture.

1.2 Dissonance Revolution

Our discussion starts with cognitive dissonance—one of the most prominent topics in social psychology. The central thesis of cognitive dissonance theory (Festinger, 1957) is that when two beliefs are inconsistent, individuals experience negatively arousing cognitive conflict (called dissonance). Because the dissonance is aversive, the individuals try to reduce it by changing one or the other beliefs. For example, when making a difficult decision, individuals show attitude change that justifies the decision. In this case, individuals who face such a decision are conflicted because not all beliefs are consistent with the decision. For example, they may have beliefs favoring the option that is rejected. The individuals are therefore motivated to reduce the conflict by justifying the decision they have made. The justification is typically achieved by changing their attitudes and beliefs so that the new attitudes and beliefs are consistent with and justify the decision that has been made. Notably, the resulting attitude change can be long lasting (Sharot, Fleming, Yu, Koster, & Dolan, 2012). By nature, then, we may be rationalizing beings, ready to justify what we have done after the fact.

Dissonance theory revolutionized social psychology by emphasizing the role of cognition in social behavior. More importantly, it also provided the first testable framework in which to conceptualize how cognition could be motivated and how the motivated cognition could yield some intriguing forms of social behavior. The theory enabled us, both in and outside of social psychology, to reflect on potentially unflattering aspects of the human mind. Indeed, the influence of dissonance theory went far beyond the field of social psychology. The term dissonance has since become incorporated into the English vernacular.

1.3 Charting the Terrain

As a scientific hypothesis, dissonance theory has been tested with three primary experimental paradigms (Aronson, 1969). First, a free-choice dissonance paradigm tests the degree to which choice leads to attitude change that justifies the choice (Brehm, 1956). By definition, difficult choices involve
competing choice options that are almost equally attractive and thus difficult to choose between. The more difficult the choice, the greater dissonance would be expected to be. To reduce the resulting dissonance, the chooser will increase her preference for chosen options and decrease her preference for rejected options. Researchers have also used an effort justification paradigm. When people invest considerable effort to obtain a positive outcome (e.g., climbing a mountain), they supposedly experience dissonance because knowing that one worked so hard is inconsistent with the possibility that the work (e.g., climbing) is valueless (Aronson & Mills, 1959). The individuals typically justify their effort by increasing their commitment to the task.

Third, in an induced compliance paradigm, individuals are led to commit an action that contradicts their beliefs or preexisting attitudes. The action therefore produces a conflict with their preexisting attitudes and beliefs. In order to reduce the resulting dissonance, the individuals change their attitudes and beliefs so that they are better aligned with their action (Festinger & Carlsmith, 1959). These three paradigms (free choice, effort justification, and induced compliance) account for the bulk of dissonance research conducted over the last half century.

Researchers have extended Festinger’s original formulation by elaborating on different theoretical possibilities (Harmon-Jones & Mills, 1999). This effort was motivated by the need to account for some important boundary conditions and moderating variables that were uncovered through extensive research over the decades. An emerging consensus is that a justification effect is typically magnified when a decision entails aversive consequences for someone else (as when fellow students may suffer from one’s decision to endorse a tuition increase) (Cooper & Fazio, 1984), or when action orientation is induced (Harmon-Jones, Amodio, & Harmon-Jones, 2009). Conversely, the justification effect is often mitigated when one’s self has been affirmed (Steele, 1988). Moreover, recent cross-cultural research shows that conditions in which a justification effect occurs vary across cultures. Specifically, whereas European Americans show a justification effect when a decision is personal and private, Asians show the effect primarily when the decision is social and public (Kitayama, Snibbe, Markus, & Suzuki, 2004).

One novel development in recent years comes from neuroscience. Several published studies have addressed brain mechanisms involved in dissonance (Harmon-Jones, Harmon-Jones, Fearn, Sigelman, & Johnson, 2008; Jarcho, Berkman, & Lieberman, 2011; Kitayama, Chua, Tompson, & Han, 2013; van Veen, Krug, Schooler, & Carter, 2009). This work has shown that dissonance may be based on a network of various brain functions such as conflict detection, reward processing, self-referential processing, and
self-regulation, among others. But implications of this newly emerging evidence have yet to be fully articulated and evaluated. In particular, it is not clear how key social psychological constructs in the dissonance literature such as dissonance, attitude or preference, effort, and justification may be mapped onto, and redefined in terms of, known brain mechanisms. Nor is it clear whether there might emerge any novel insights or testable implications through this effort to reformulate dissonance from the neuroscience perspective.

Also important is an effort to explore dissonance among nonhuman animals. Since pioneering work by Lawrence and Festinger (1962), dissonance effects have been repeatedly demonstrated with nonhuman animals, including rodents (Lydall, Gilmour, & Dwyer, 2010), monkeys (Egan, Bloom, & Santos, 2010), pigeons (Clement, Feltus, Kaiser, & Zentall, 2010), and starlings (Kacelnik & Marsh, 2002). All these nonhuman animals are not capable of using language and, thus, supposedly also deprived of any higher-order beliefs. This animal literature therefore calls into question all theoretical accounts, including the original theory by Festinger (1957) and its reformulation in terms of self-perception processes (Bem, 1967), that posit sophisticated cognitive beliefs as necessary elements of dissonance. It is possible that cognitive beliefs often participate in the process of dissonance arousal and reduction (Stone & Cooper, 2001). However, such beliefs may be neither necessary nor sufficient. The fact that humans with little or no episodic memory capacity show a full-fledged dissonance effect (Lieberman, Ochsner, Gilbert, & Schacter, 2001) also underscores the secondary role of cognition since episodic memory is supposedly required to cognitively process prior experience that feeds into dissonance (e.g., having chosen one option over the other). There may be something else, other than cognitive beliefs, that defines the core of dissonance, but this “something else” has yet to be clearly defined and explicated.

We should bear in mind there is one curious omission from previous theorizing on dissonance. As noted earlier, although dissonance theory has spawned hundreds, if not thousands, of experiments, all this work focuses on how the decision maker might resolve a cognitive conflict posed by the decision she has made. The reason for this may in part be due to an insistence of Festinger (1957) that for dissonance to arise, a decision must be final and irreversible. It is possible that an irreversible decision entails a strong commitment to it and, most likely, an equally strong tendency to act on it (Lewin, 1947); so all information contradicting it could be a cause of a much larger conflict. Thus, it is possible that the finality or reversibility of a decision is an important contributing factor to dissonance (Knox & Inkster, 1968). However, it is
debatable whether no behavioral conflict is involved when decisions are not final. In fact, we propose that behavioral conflict can be quite intense and consequential even before any decisions are made.

Because of its emphasis on postdecisional processes, dissonance theory is agnostic about how the decision is made. To fill this important gap, subsequent researchers have proposed cognitively elaborate decision mechanisms that offer insights into dissonance effects (Shultz & Lepper, 1996; Simon, Krawczyk, & Holyoak, 2004). However, these models minimize the significance of affect or motivation. Moreover, they are hard pressed to account for the simple fact that seemingly identical effects are present in nonhuman animals that are cognitively far less equipped than human decision makers. In short, it is unknown whether postdecisional dissonance might be related to preceding decision mechanisms and, if so, how the two mechanisms might be related. In the current paper, we propose that the same conflict resolution mechanisms may be implicated in how decisions are made, as well as how they are justified once they have been made.

Last, but not least, one important strength of dissonance theory is its breadth. Ironically, however, after the theory lost its initial traction in the 1980s and 1990s, much of the literature on motivation has proceeded without considering potential implications of dissonance-related mechanisms. Some of the affective and motivational phenomena that have been investigated without consideration of dissonance processes include behaviors of nonhuman animals that appear to show a type of “work ethic” (Clement et al., 2010), ego-depletion (Baumeister et al., 2007), and addiction (Flagel et al., 2010). It remains to be seen how such diverse phenomena might be integrated within a coherent theoretical framework.

# 2. THE BIOSOCIAL MODEL

## 2.1 Key Propositions

To address the gaps in the literature noted earlier, we propose a new model of affective decision making called the biosocial model. As illustrated in Figure 1, the model is composed of four core propositions: negatively arousing behavioral conflict, active search for positive incentives, recursive loop, and top-down/bottom-down regulation.

### 2.1.1 Negatively Arousing Behavioral Conflict

We assume that difficult decisions and demanding tasks cause a behavioral conflict, which is negatively arousing. This complex of both behavioral
conflict and negative arousal constitutes the core of dissonance as proposed by Festinger. Whereas the original theory suggested that dissonance by definition was confined to conflict that is generated by a decision and, thus, the dissonance is construed to be postdecisional, the current model emphasizes that dissonance is inherent in the decision itself. As we shall see, the model hypothesizes that the same dissonance mechanism can be recursively engaged. It therefore can be extended to cover postdecisional dissonance phenomena.

A behavioral conflict emerges once a decision is required to select between mutually incompatible responses or when a task is too demanding that there arises a temptation to stop and quit. At the most primitive level, this conflict is apparent when nonhuman animals seek a reward ("go" response) that can be attained only by overcoming certain difficulties ("no-go" response). Of course, in many cases involving human adults, the conflict can be more elaborate and informed by cognitive beliefs. For example, a similar conflict is apparent when one is induced by an

Figure 1 The biosocial model of affective decision making. Demanding tasks and difficult decisions induce behavioral conflict, which is negatively arousing. To reduce the negatively arousing conflict, available behavioral options are parsed for unique positive incentives. Once identified, such incentives resolve the conflict, tame the negative arousal, and enable one to select a response. This decision mechanism may be recursively engaged. The conflict detection system is regulated by both top-down and bottom-up neural pathways. dACC and aINS play significant roles in detecting negatively arousing behavioral conflict. Incentive processing is likely carried out at reward-processing areas including vSTR and om/vmPFC. Top-down regulation of dACC (behavioral conflict detection) is likely subserved by left dlPFC, whereas the amygdala and the midbrain DA system are likely implicated in the bottom-up regulation of dACC.
experimenter to agree with a counter-attitudinal position or when one is choosing between two equally attractive vacation plans. However, no matter how elaborate the background beliefs might be, the resulting conflict is constituted at the level of behavioral response, whether overt (e.g., approach vs. avoidance) or covert (e.g., approval vs. disapproval). The biosocial model therefore assumes that the core of dissonance lies in this conflict between different behavioral representations (Zajonc & Markus, 1984).

2.1.2 Search for Positive Incentives

Consistent with the original dissonance theory, the biosocial model assumes that behavioral conflict is negatively arousing. As a consequence, the decision maker, both human and nonhuman animals alike, is motivated to reduce it. Unlike the original dissonance theory, the biosocial model holds that resolving behavioral conflict, rather than resolving cognitive inconsistency, is the core of dissonance reduction. It is hypothesized that the negatively arousing behavioral conflict is best reduced by identifying positive, appetitive incentives in one of the response options. By identifying such incentives, the decision maker will be able to reduce the negative arousal while resolving the conflict. Importantly, this affective information produces a potent appetitive response tendency, which would enable the decision maker to either achieve a clear-cut decision if this process is engaged before the decision or reinforce the decision-consistent attitude if it is engaged afterward. It is worth emphasizing that implications of the basic principle proposed here, namely, that of behavioral conflict leading to an active search for positive incentives, goes beyond dissonance itself. As we shall see, it covers various motivational phenomena, including sunk-cost fallacy, addiction, and ego-depletion. It further sheds a new light on how and why humans may become motivationally and affectively attached to religion, tradition, and other forms of cultural symbol.

Much of the process involved in both arousal and reduction of dissonance can be entirely precognitive. That is, positive, appetitive information that is identified during a decision may be no more than a transitory activation of subcortical reward-processing regions of the brain (Berridge & Robinson, 1995; Zajonc, 1980). It should be noted, however, that especially in the case of human adults, the precognitive affective information may be cognitively encoded and elaborated to form the basis of a new full-fledged attitude. The initial affect that is produced to resolve the dissonance promotes a certain behavioral response while at the same time yielding a congruent attitude. Thus, the resulting attitude will inevitably be suitable
for justifying the decision or response that is being developed. As implied by the original dissonance theory, the biosocial model predicts that post-decisional justification is likely to be quite common and powerful. However, unlike the original dissonance theory, the biosocial model holds that the “affective seed” for justification is formed during the decision and plays a critical role in informing subsequent cognitive rationalizations.

2.1.3 Recursive Loop

According to the biosocial model, the decision mechanism specified in terms of the first three propositions is sometimes (but not always) recursively engaged to magnify the initial decision-based attitude change. This recursion may account for certain important phenomena that involve post-decisional processes (Cooper & Fazio, 1984). For example, people rationalize their decisions more if they learn that the decisions are irreversible (Knox & Inkster, 1968). The current model suggests that once a decision is made, it by itself can be a source of a new behavioral conflict, which in turn will initiate another round of a search for positive incentives. This conflict may be larger if the initial decision is irreversible, thereby producing a stronger response tendency that may be pitted against any competing response tendencies.

Moreover, through the recursion, both existing conflicts and associated negative emotional arousal are likely to be consciously registered and subjectively experienced. As a consequence, these experiences themselves may become the target of conscious reflection and interpretation. One important consequence of this is that the negative arousal evoked by behavioral conflict may be interpreted in terms of other events in the environment. For example, the negative arousal may be misattributed to external events. When this misattribution happens, the effort to reduce the negative arousal may dissipate (Cooper & Fazio, 1984; Zanna & Cooper, 1974).

2.1.4 Top-Down/Bottom-Up Regulation

The biosocial model holds that the magnitude of behavioral conflict (i.e., dissonance) is modulated by a variety of situational factors. For example, imagine someone who is driving a car on a busy street. In such situations, the person will be more careful than usual. To put it differently, she will recruit self-regulatory regions of her brain to “tighten the belt” of the conflict detection system (which is required to detect an error so as to preempt any serious accidents). She is thus using her higher-order goal of “driving carefully” to regulate her conflict detection system. This pathway may
therefore be called top-down. When the conflict detection system is upregulated, behavioral conflicts will be more likely to be taken note of and thus can initiate a more vigorous search for positive incentives. Conversely, if the system is downregulated (as when the driver is relaxed), the conflicts may not be noticed even if they exist, thus entailing no downstream effects including the search for positive incentives.

An analogous regulatory process may happen bottom-up. For example, the same person may be exposed to a cue indicating certain threats impinging on her (e.g., a police car passing her car at high speed). Such a cue will promptly increase the level of alertness and, as a consequence, enhance the sensitivity of the conflict detection system. In this case, it is the immediate emotional reaction of fear that is regulating the conflict detection system. Thus, the mechanism is said to be bottom-up. As we shall see, both top-down and bottom-up regulatory pathways for the sensitivity of the conflict detection system are important in understanding an assortment of variables that have been proposed as moderators of dissonance effects, including action orientation, self-affirmation, mortality salience, and culture.

2.2 Moving Forward

With this overview in mind, we now turn to each key component of the biosocial model in some detail. This review is divided into several major sections. To begin, we will discuss the core proposed mechanism of affective decision making, composed of negatively arousing behavioral conflict followed by a search for positive incentives in terms of key brain substrates that support it (Section 3). We will show that the same mechanism is responsible, not only for phenomena traditionally subsumed under the rubric of cognitive dissonance, but also an assortment of disparate phenomena including animal behavior that mimics cognitive dissonance, addiction, and ego-depletion. Next, we will discuss neural mechanisms involved in both the top-down (Section 4) and the bottom-up regulation of dissonance (Section 5). Finally, we will discuss the nature of a recursive loop involved in affective decision making, which has implications for postdecisional effects such as effects of aversive consequences and misattribution (Section 6). Taken together, we will show that the biosocial model integrates insights from both dissonance theory as originally formulated and all subsequent elaborations. Moreover, the model offers important new insights on apparently unrelated motivational phenomena. One strength of the model is to provide a single framework that is well grounded in
current knowledge of how the brain functions. This framework will be applicable to a wide range of affective and motivational effects. In the concluding section, we will explicate additional implications of the model and suggest directions for future research on biosocial processes involved in dissonance in particular and affective decision making in general.

3. NEURAL SUBSTRATES OF AFFECTIVE DECISION MAKING

3.1 Behavioral Conflict and Negative Arousal

Numerous studies in cognitive neuroscience demonstrate the critical role of the dorsal anterior cingulate cortex (dACC; illustrated in Figure 2) in the detection of behavioral conflict in a variety of conflict-inducing paradigms such as Stroop tasks and flanker tasks (Botvinick, Cohen, & Carter, 2004; Bush, Luu, & Posner, 2000; Carter & van Veen, 2007; Shenhav, Straccia, Cohen, & Botvinick, 2014; but also see Izuma & Adolph, 2013). An event-related potential (ERP) component marking the detection of behavioral conflict, called error-related negativity (ERN), is thought to be derived from dACC (Carter & van Veen, 2007), although the issue is far from settled (Agam et al., 2011). It stands to reason that behavioral conflicts associated with demanding tasks or important decisions are also detected by the dACC. Consistent with this analysis, several researchers have proposed that dACC

Figure 2 Some key brain regions involved in arousal and reduction of dissonance as specified in the biosocial model. (A) Dorsal anterior cingulate cortex (dACC). (B) Caudate nuclei. (C) Orbital/ventral medial prefrontal cortex (om/vmPFC). (D) Ventral striatum/nucleus accumbens (vSTR/Nacc). (E) Anterior insula (aINS).
constitutes a key neural substrate of dissonance (Harmon-Jones & Harmon-Jones, 2008; Kitayama et al., 2013; van Veen et al., 2009).

dACC is also important in detecting reward prediction errors. When individuals anticipate a certain reward outcome (e.g., earning $5 in a gamble), there arises a negative reward prediction error if they fail to get the outcome. Conversely, if the outcome is better than anticipated, there arises a positive reward prediction error (Walsh & Anderson, 2012). A negative reward prediction error signals the outcome being less than expected. This error signal may interrupt and suspend an approach tendency to achieve the reward and, as such, lends itself to a go versus no-go behavioral conflict. For the purpose of the current discussion, then, we hypothesize that negative reward prediction error can be subsumed under a larger category of behavioral conflict.

Anatomically, dACC is densely connected with the anterior part of the insula cortex (anterior insula or aINS) (see Figure 2). aINS is thought to map visceral bodily sensations (Damasio & Carvalho, 2013; Immordino-Yang, Yang, & Damasio, 2014; Singer, Critchley, & Preuschoff, 2009), which may then be cognitively interpreted to yield subjective feelings (Immordino-Yang et al., 2014; Schachter, 1964). In particular, aINS shows strong activation during the experience of certain negative states or emotions such as pain and disgust (Damasio & Carvalho, 2013). The network including dACC and aINS is activated during the experience of social pain that is caused by social exclusion (Eisenberger & Lieberman, 2004). This network also responds to vicarious pain or empathy for in-group members (Han et al., 2009; Singer, 2004; Xu, Zuo, Wang, & Han, 2009). The network including dACC and aINS is therefore characterized as a “neural alarm system” (Eisenberger & Lieberman, 2004). Further, both dACC and aINS are correlated with autonomic arousal, which is known to occur when dissonance is induced (Croyle & Cooper, 1983; Elkin & Leippe, 1986; Losch & Cacioppo, 1990).

Altogether, we may hypothesize that when trying to make a difficult decision or working on a demanding task, dACC detects behavioral conflict, and this activation quickly spreads to aINS to make the conflict negatively

---

1 Existing research on animal behavior suggests that ACC also encodes reward values of different actions or incentive cues. Thus, laboratory animals with permanent or transient ACC lesions cannot perform optimally when seeking rewards (Amiez, Joseph, & Procyk, 2006; Hillman & Bilkey, 2012; Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006). The function of ACC to encode reward values makes this region suitable for the detection of reward-related conflicts, such as when one has to engage in a costly action (e.g., effort or investment) to achieve a certain desirable goal.
arousing. This negative arousal shares much in common with experiences of social pain. Moreover, aINS is known to respond to disappointment and regret (Chua, Gonzalez, Taylor, Welsh, & Liberzon, 2009). We may thus suggest that the complex encompassing dACC and aINS constitutes the core of negatively arousing behavioral conflict, or dissonance as formulated by Festinger (1957). All subsequent dissonance researchers have conceptualized dissonance as an aversive state (Cooper & Fazio, 1984; Harmon-Jones et al., 2009; Steele, 1988) that has a negatively arousing motivational potential (Elliot & Devine, 1994). The current model adds to the previous conclusion by proposing that the network of dACC and aINS supports this aversive motivational state.

Evidence for the proposal above comes from a recent neuroimaging study (Kitayama et al., 2013), in which American undergraduates were scanned while making choices between many pairs of popular music CDs. They had been told that one of the chosen CDs would be given to them as a gift at the end of the session. To investigate the effect of choice difficulty on brain activity, half of the CD pairs were composed of two CDs that had been rated by each participant to be nearly equally likable, whereas the remaining half consisted of two CDs that had been rated to be very dissimilar in likability. In support of the hypothesis that dissonance entails strong activation of the network consisting of both ACC and aINS, this

---

2 Chen and Risen argue that a certain statistical artifact may be involved in a justification effect observed in the free-choice paradigm (Chen & Risen, 2010). Specifically, in this paradigm, individuals are typically asked to rate a limited set of options. They are then given a choice between two options that are similarly rated. The same rating task is administered again after the choice. It is assumed that the prechoice rating is error-free. In reality, however, there may be measurement error such that even though the two options appear very close in liking in the first rating task, their true likability might in fact be further away from one another. In this case, the truly more favored option may be both selected and ranked higher in the subsequent measurement. This stochastic mechanism, if sufficiently powerful, will ensure what appears to be a justification effect. There is a debate regarding how serious this potential artifact might be (Izuma & Murayama, 2013; Kitayama, Tompson, & Chua, 2014). We should bear in mind, however, that especially when measurement error is minimized with careful experimental control, the artifact is likely to be negligible. We pointed out elsewhere that if such an artifact is always involved, there should emerge a sizable effect of apparent justification even when dissonance is predicted to be minimal (Kimel, Lopez-Duran, & Kitayama, 2014). As shown in numerous studies—especially behavioral studies involving only one choice (which are reviewed later Sections 5.3 and 5.4), a justification effect that is observed in one condition of a given study disappears in another condition of the same study, depending on dissonance-increasing or dissonance-depressing experimental manipulations. The absence of any justification effect in at least one condition within a study shows that measurement error is negligible within the study under discussion. Hence, a justification effect observed in all other conditions of the study can hardly be due to any statistical artifact.
study observed strong activations in these two areas (illustrated in Figure 3) during difficult (as compared to easy) choices. Another recent study tested Japanese participants and also found that choice difficulty was positively correlated with dACC activation during choice (Izuma et al., 2013).

In another fMRI study, van Veen and colleagues (2009) adopted an induced compliance paradigm (Cooper & Fazio, 1984). American college undergraduates first worked on an extremely boring task in the unpleasant scanner environment for 40 min. Afterward, they were asked to indicate whether they would endorse a series of sentences by pressing a response key. Some of the sentences focused on the first phase of the study (I feel calm and peaceful in the scanner), whereas the rest were irrelevant to the latter (The weather is very bad today). They were asked to respond as if they enjoyed the first phase of the study. Half of the participants were told that every time they indicated this attitude they would receive one dollar (control condition). The remaining half were told that in the scanner control room a patient scheduled to be scanned after the participants was waiting. The person was so anxious and worried, so the participants were told that the experimenter would like them to respond as if they enjoyed the study. In that way, the patient would feel at ease. They were told that the patient would be watching the participants’ responses in the control room (dissonance condition). Because one’s own action can have unintended, potentially negative effects on the patient, the decision to act that way should produce conflict between conforming to the request from the experimenter and resisting the pressure so as to avoid negative effects on the patient. As may be expected, van Veen and colleagues observed an increased activation

Figure 3 Areas of the brain that are activated by difficult choices, relative to easy choices in a free-choice dissonance study by Kitayama et al. (2013). Reprinted from Kitayama et al. (2013) with permission from Elsevier.
of both dACC and aINS, while the participants were reading the critical sentences relative to reading the control sentences in the dissonance condition. Their primary results are shown in Figure 4.

### 3.2 Making Decisions: Reward Processing and Response Selection

Once negatively arousing behavioral conflict (i.e., dissonance) is aroused, it will motivate a search for positive incentives that are unique to one of the available response options. For example, when choosing between a rock CD and a modern jazz CD, the decision maker might think of his plan to host a party next week and imagine how much better the rock music would be for the occasion than the jazz music. For another example, think about a climber, who saw a ridge in front of him and anticipated arriving at the summit only to realize that he still had a long way to go. Every time this happens, his goal to reach the summit is interrupted. This interruption, caused by a negative prediction error indicating that the outcome (just another ridge) was worse than expected (summit), results in a conflict between the two

**Figure 4** Areas of the brain that are activated by target sentences (vs. control sentences) in the dissonance condition as compared to the control condition in an induced compliance study performed by van Veen et al. (2009). The dorsal anterior cingulate cortex (dACC) and the anterior insula (aINS) in combination constitute the core of dissonance. The activation of the midbrain area may indicate successful effort to find positive features of the preceding, seemingly boring task (see Section 3.3, Page 92). Reprinted from van Veen et al. (2009) with permission from Macmillan Publishers Ltd., copyright 2009.
behaviors (i.e., continuing to climb vs. quitting), which would prompt him to look for added incentives in the endeavor. Once found and recognized, the additional incentives would resolve the conflict and reduce the negative arousal. Moreover, the newly identified incentive will boost the affective value of the chosen option. Thus, the climber will be even more motivated to reach the mountaintop as a result of having experienced the behavioral conflict.

The link between negatively arousing behavioral conflict and the search for positive incentives may be biologically hard-wired. Such a biological mechanism may make sense from an adaptive point of view. The production of positive affect to neutralize negative affect is instrumental in maintaining homeostasis (Solomon & Corbit, 1974). Moreover, as pointed out by some animal researchers, the mechanism to boost motivation under repeated failures to obtain reward may be highly adaptive for foraging animals (Clement et al., 2010). In fact, uncertainty in reward anticipation is known to increase the appetitive conditioning to cues associated with the reward (Anselme, Robinson, & Berridge, 2013), hence increasing the motivation to seek the reward. The same reasoning could undoubtedly apply to hunters and gatherers and perhaps to modern humans who face fierce competition in business and other professional circles, although this latter point remains speculative. The mechanism of searching for positive incentives in goals that are blocked may also be learned. As some researchers point out (Harmon-Jones et al., 2009), action may typically be more adaptive than inaction as a strategy to deal with behavioral conflict. If so, most people are rewarded more by searching for positive incentives in pertinent goals. They may therefore acquire this mechanism as an effective way to handle behavioral conflict. For our purposes, the origin of the mechanism is less important than the mechanism itself.

During the active search for positive incentives in one response option to choose and pursue, there will be coordination between memory search for relevant previous experience or prospective plans and attention directed to one of the options. In all likelihood, retrieval of some initial bit of information that makes one of the options slightly more attractive would foster a shift of attention toward the option, which in turn would encourage further search of relevant information in memory. This close interaction between memory search and attention will lead to the identification of additional incentives in the option (Shimojo, Simion, Shimojo, & Scheier, 2003; Simion & Shimojo, 2006). Once these positive incentives are identified
in one of the decision options, they will allow a choice to be made, thereby reducing the initial behavioral conflict.

Consistent with the hypothesis that behavioral conflict motivates an active search for incentives linked to a response option to select and pursue, previous work shows that when one’s goal is thwarted, the person is motivated to more vigorously pursue the goal (Brehm, 1966). More generally, people may become more engaged in a goal when it is blocked (Higgins, 2006). These effects are likely to be mediated by the search for incentives for blocked goals. Another analogous process has been discussed by opponent process theory (Solomon & Corbit, 1974). The theory argues that when negative affect (e.g., dissonance) occurs, it automatically recruits a slow-growing positive affect so as to maintain affective homeostasis. The biosocial model is consistent with opponent process theory insofar as both theories assume that the initial negative affect is sufficient to build up positive affect. Whereas opponent process theory merely postulates this contingency, the biosocial model assumes that it is mediated by an active search for positive incentives. Unlike the biosocial model, opponent process theory further assumes an analogous process for positive affect, insisting that initial positive affect is sufficient to produce negative affect that is suited to maintain homeostasis. This premise of opponent process theory may be questionable, insofar as, contrary to the theory, people often savor positive affect (Speer, Bhanji, & Delgado, 2014) and even try to build upon it (Fredrickson & Losada, 2005). Altogether, there is a solid basis to hypothesize that once negatively arousing behavioral conflict (dissonance as implemented in dACC and aINS) is activated, it automatically fosters an active search for positive incentives linked to a response option to select and pursue.

Specific mechanisms involved in identification of positive incentives and subsequent response selection (i.e., decision) are illustrated in Figure 5. The figure also notes the specific brain anatomy that supports these decision mechanisms. Researchers have suggested that the ventral striatum (vSTR), including the nucleus accumbens (Nacc), is particularly responsive to cues previously associated with rewards (i.e., incentives) (see Figure 2; Knutson & Cooper, 2005; Shenhav & Buckner, 2014). When such incentive cues are identified in one of the available options (e.g., the rock CD is very suitable for the upcoming occasion, or the growing expectation of a fabulous view from the summit), they will cause an increase of activity of dopaminergic (DA) neurons in the midbrain area (Bromberg-Martin & Hikosaka, 2009; Watanabe, Lauwereyns, & Hikosaka, 2003). This midbrain
DA activation will then spread to the striatal reward-processing region (vSTR/Nacc), where the incentive information is further processed. At this stage of processing, the activation of the reward-processing area is only transitory and, by itself, offers neither any motor plan (approach behavior toward the source of the positive incentive) nor any cognitive representations of the incentives. These more elaborate operations are carried out in subsequent processing steps, which involve both bottom-up and top-down pathways through which the incentive information is used to control response selection (Frank & Claus, 2006).

At this stage of processing, the activation of the reward-processing area is only transitory and, by itself, offers neither any motor plan (approach behavior toward the source of the positive incentive) nor any cognitive representations of the incentives. These more elaborate operations are carried out in subsequent processing steps, which involve both bottom-up and top-down pathways through which the incentive information is used to control response selection (Frank & Claus, 2006).

First, the incentive information may spread bottom-up through the caudate nuclei to the premotor cortex (Figure 2). Through this pathway, DA activation will then spread to the striatal reward-processing region (vSTR/Nacc), where the incentive information is further processed.³

Although DA neurons are linked closely to reward processing, it is a mistake to confine them only to reward processing (Salamone & Correa, 2012). For example, the DA reward function highlighted in the current discussion is likely to coexist hand in hand with DA’s role in enhancing glucocorticoid receptors in the prefrontal cortex, which are linked to “stress” responses (Butts, Weinberg, Young, & Phillips, 2011). It is therefore possible that dissonance produces a stress, which in turn increases prefrontal DA concentrations. This pathway may also have downstream consequences such as compromised executive functions. This pathway may be relatively slow-growing and, thus, may prove to be more relevant in cases involving dissonance-related distress of the sort that is long lasting as when individuals are torn apart over days or weeks over the decision they made. These cases are beyond the scope of the current discussion. However, they are obviously important and should be addressed in future work.
the response is directly controlled by the incentive information. Selection of
the stimulus that carries the positive incentives that have been identified will
be strongly primed. Second, it may also be cognitively represented, thus
likely experienced subjectively (Brosch & Sander, 2013). The key site to
cognitively represent positive incentives includes the orbital/ventral medial
prefrontal cortex (om/vmPFC) (Elliott, Dolan, & Frith, 2000; Figure 2).
A cognitive representation of the incentives that emerges may be combined
with other, more context-specific cognitions about the impinging social sit-
uation. The resulting contextualized understanding of the incentive will be
used to direct the response selection (Frank & Claus, 2006). For example,
the response tendency toward the option with the identified incentives
(generated through the bottom–up pathway involving the caudate) may
be tempered or facilitated, depending on how appropriate such an action
might be given the specific context in which the decision is being made.

3.3 Testing the Decision Mechanism
Evidence for the decision mechanism above comes from the Kitayama et al.
(2013) study discussed earlier and another similar study by Jarcho et al.
(2011). In both studies, participants rated decision options (CDs in Kitayama
et al., and names and paintings in Jarcho et al.) both before and after choosing
one option from each of many pairs of options. Kitayama et al. observed
that after the choices, clear choice justification was observed. That is, right
before the choices, there were no differences in liking for CDs that were to
be chosen and those that were to be rejected. However, right after the
choices, the chosen CDs were rated as significantly more likable than the
rejected CDs. In the Jarcho et al. (2011) study, the researchers only used
data from approximately 60% of participants who showed a reliable choice
justification effect.

Importantly, both studies scanned participants while they made choices
and tested whether any brain areas showed in-choice activations that
predicted a greater justification effect. The two studies revealed two con-
verging findings. First, in both studies, in-choice activity in a part of the sub-
cortical reward-processing area (the vSTR, including the Nacc), was reliably
related to postchoice preference change. The particular region identified in
the Kitayama et al. (2013) study, shown in Figure 6A, was Nacc, which is
part of the vSTR (Knutson & Cooper, 2005). Jarcho et al. (2011) also found
the in–choice activity in the vicinity of vSTR predicted subsequent prefer-
ence change. As noted previously, these regions are strongly associated
with anticipated reward and, thus, are thought to show incentive processing (Berridge, Robinson, & Aldridge, 2009; Knutson & Cooper, 2005; O’Doherty, Deichmann, Critchley, & Dolan, 2002; Shenhav & Buckner, 2014). Kitayama et al. interpreted this finding as indicating that on some choice trials, the decision maker successfully identified positive features, or incentives, associated with one of the options, which led to an increased preference for the option and, thus, allowed her to choose it. When the increased preference was assessed at a later point, the preference was correlated with the preceding Nacc activity. This evidence is consistent with the claim of the biosocial model that decision-consistent affective information is already activated at the precognitive level during choice, before a choice is actually made. It is this affective information that would enable one to make the choice. This information is generated so as to reduce behavioral conflict and used to guide the choice.

Both Jarcho et al. (2011) and Kitayama et al. (2013) also found that postchoice preference change was predicted by in-choice activity of the posterior cingulate cortex and adjacent precuneus (PCC/Pcu). PCC/Pcu is linked to autobiographic memory (van der Meer, Costafreda, & Aleman, 2010). Thus, this evidence is consistent with the hypothesis that the decision maker searches in memory for information that makes certain features more positive. Often, this information is self-relevant and thus episodic, as in the example above where the decision maker thinks about his plan to host a party next week and interprets one of the CDs in light of this plan. Thus, the more intensely memory is searched for such information (as shown

![Figure 6](image-url) In-choice activation of the two areas of the brain predicted postchoice increase of preference for chosen options in the free-choice fMRI study by Kitayama et al. (2013). (A) Nucleus accumbens (Nacc) and (B) posterior cingulate cortex (PCC). Reprinted from Kitayama et al. (2013) with permission from Elsevier.
by the PCC/Pcu activation), the more likely it will be for the decision maker to find a positive distinctive feature that would enable her to make a clear choice.4

As discussed previously, in the fMRI study on induced compliance by van Veen and colleagues (2009), while reading and responding to the stimulus sentences relevant to the first “boring phase” of the study, participants in the dissonance condition showed increased activity in both dACC and aINS. Importantly, in this study, the participants in the dissonance condition also showed increased activation in the midbrain area (see Figure 4). Because the midbrain area has high concentration of dopamine neurons, which respond to reward processing, it is possible that this activation signaled the hypothesized search for positive, rewarding experiences in the seemingly “boring” study the participants completed.

Three additional fMRI studies have utilized a free-choice paradigm similar to the one by Kitayama et al. (2013) and Jarcho et al. (2011), but examined neural activation during the prechoice and postchoice rating tasks, instead of in-choice neural activation. First, Izuma and colleagues (2010) had Japanese participants complete a similar free-choice paradigm with food options (Izuma et al., 2010). The researchers found a significant postchoice decrease of liking for rejected food options, but there was no postchoice increase of liking for chosen food options. This postchoice decrease of liking for rejected food options was mirrored in a corresponding decrease of vSTR activation for the rejected food options. Again, there was no effect for the chosen food options. Unfortunately, this study suffered from one idiosyncratic procedural aspect: Participants were explicitly reminded whether they had chosen each food option when they rated it for the second time after the choice. The results might therefore have reflected responses to this reminder rather than any representations of the pertinent food options that might be altered by the choice. The same confound existed in another imaging study by Qin et al. (2011). Hence, for our purposes, implications of these studies are uncertain.

4 In the Kitayama et al. study, the Nacc and the PCC/Pcu were the only two areas that were related to postchoice preference change. In the Jarcho et al. study, however, two additional areas showed similar relations with postchoice preference change: the medial prefrontal cortex (mPFC) and the inferior frontal gyrus (IFG). Some caution is due, because the findings were not replicated in the Kitayama et al. study. Nevertheless, it is instructive that mPFC is also involved in self-relevant information processing, including episodic memory (Addis, Wong, & Schacter, 2007). Furthermore, IFG is often involved in emotion regulation. As Jarcho et al. argued, it is possible that IFG was implicated in downregulating negative affect associated with difficult decisions. In support of this analysis, Jarcho et al. found a reliable connectivity between IFG and aINS, such that strong IFG activity was linked to attenuated aINS activity.
Fortunately, the Sharot, De Martino, & Dolan (2009) study did not have any of such interpretive problems. The researchers asked British participants to rate how much they would enjoy many vacation destinations, make hypothetical choices between two vacation destinations, then rate the vacation destinations again. The researchers observed that preference ratings for chosen options increased, and those for rejected options decreased, after the choice. Moreover, they observed a postchoice increase in activation of the caudate nucleus for chosen (vs. rejected) options. The caudate nucleus is part of the brain area serving reward processing and action control, located between the vSTR and the premotor cortex, above vSTR, but below the premotor cortex (see Figure 2). It is thought to relay the reward information processed in vSTR to the premotor cortex. Moreover, this post-choice increase of caudate activation for chosen vacation destinations correlated with a comparable increase of self-reported preferences of the destinations. In light of the processing model in Figure 5, we may assume that affective information identified during choice (a function supposedly carried out in vSTR/Nacc) was transferred upward to the caudate region so as to control motor output, both actuarial and imaginary.

3.4 Effort Justification

One typical decision conflict occurs when a rewarding action is costly, as when arriving at a mountaintop on foot (rather than, say, transported to the top by a helicopter). The climber will be eager to reach every ridge, only to realize that there is (at least) one more to go before finally arriving at the mountaintop. This negative “prediction error” (the signal indicating the outcome is worse than expected) discourages the climber, thus providing a potent source of conflict with the effort he has expended to make it to the top. Another example involves initiation rituals that are required for one to join a special social club (Aronson & Mills, 1959; Atkinson & Whitehouse, 2011; Gerard & Mathewson, 1966). The attractive goal (joining the club) initiates an approach tendency, which is often interrupted and disrupted by difficulties and obstacles (hard, often painful, work), thereby resulting in a strong behavioral conflict. According to the biosocial model, when facing such a conflict, dACC is activated, which entails an activation of aINS. In combination, activation of this dissonance complex (dACC + aINS) prompts a search for additional positive incentives associated with the rewarding goal that is being pursued (reaching the mountaintop or joining the social club in the examples above). The incentive value of the goal should become even greater than before. This is exactly what a classic dissonance study by
Aronson and Mills (1959) found. Individuals showed elevated levels of commitment to a social club when they were required to go through a severe (vs. light) initiation ritual in order to join the club. This phenomenon, traditionally called effort justification, has since been replicated (Axsom & Cooper, 1985; Gerard & Mathewson, 1966).

Aronson and Mills (1959) interpreted their effort justification finding in terms of reduction of cognitive dissonance as originally formulated by Festinger (1957). That is, a belief that one is undergoing painful action (effort) is inconsistent with another belief that joining the club is not valuable. To resolve this cognitive inconsistency, individuals supposedly changed the belief that was more amenable to change, namely the evaluation of the club itself. Thus, the actor is said to justify the effort she expended to achieve her original goal by increasing the value of the goal. According to this interpretation, however, there must be cognitive beliefs to begin with. Moreover, the evaluation of the club must be more amenable to change than the evaluation of the pain associated with the effort. Our formulation requires neither of these assumptions. It holds that there is a conflict between the go tendency associated with the attractive club and the no–go tendency associated with painful effort. Because this conflict detected at dACC is negatively arousing, due to dACC’s connections with aINS, individuals look for additional incentives in the goal at issue. For this reason, the club—the most salient incentive that is available in the situation—becomes even more rewarding and is evaluated more positively.

Effects similar to effort justification have been observed in a series of behavioral economics studies on the so-called “sunk-cost fallacy.” According to this phenomenon, once individuals invest resources (e.g., time, effort, and money) on a certain decision option, they find it extremely hard to forego the investment, even when the pursuit of the initial option is no longer profitable and thus not desirable. The failure to forego the investment is considered irrational because all it does is to lead to further loss of resources (Cunha & Caldieraro, 2009). Although this phenomenon can be mediated by sophisticated cognitive inferences (e.g., mental accounting), as typically assumed in the current judgment and decision-making literature (Thaler, 1999), we propose that it is also likely to involve a precognitive affective mechanism, such that initial investment in an option is inherently conflict-prone and, thus, motivates the decision maker to identity positive incentives in the action itself.
3.5 Animal Dissonance

Both the dissonance complex that is activated in effort justification (dACC and aINS) and the affective processing involved in the reduction of dissonance (vSTR/Nacc) are not unique to humans. In fact, these brain structures are shared with nonhuman animals including other mammals (e.g., rodents and monkeys) and birds. It should therefore not come as a surprise that these nonhuman animals exhibit analogous effects.

Much of the research on “effort justification” in animals involves requiring animals to repeat a certain set of behaviors in order to receive a food reward. For example, in one experiment, starlings were trained to repeat flying 1 m in the air from perch to perch, and once they repeated it the required number of times, they were allowed to press a key to obtain a food reward (Kacelnik & Marsh, 2002). On some trials, the animal had to fly in the air only four times before they received a food reward (low effort condition), and on other trials, they had to fly 12 times to receive the same food reward (high effort condition). The keys were color-coded such that the key was in one color (e.g., green) on the high effort trials but another color (e.g., blue) on the low effort trials. In a test trial, the birds were given a choice between the two-colored keys to obtain food. The birds were reliably more likely to choose the key associated with high effort.

In a paper aptly titled “‘Work ethic’ in pigeons,” Clement and colleagues have shown nearly identical effects (Clement et al., 2010). In the low effort condition, pigeons were trained to press a center key just one time, and then they could choose to press one of two color-coded keys. The pigeons learned that the red key yielded a food reward, whereas the yellow key yielded no food. The high effort condition was identical, except that the pigeons had to press the center key 20 times before choosing between two color-coded keys (blue and green). After the training sessions, the pigeons were given a choice between the key that yielded a food reward in the low effort condition and the key that yielded a food reward in the high effort condition. Importantly, the pigeons received the same amount of food in both conditions, so any difference in choice rates should have been due to level of effort rather than initial value of the reward. Once again, the birds preferred the keys associated with high effort.

The research on starlings and pigeons measured choices between options associated with high or low effort, but did not directly measure preference for the rewarding food itself. A recent study used rats and sought to go a step further by measuring palatability of a sucrose solution used as reward (Lydall...
et al., 2010). To measure palatability, the researchers performed a micro-
structural analysis of licking during ingestion. Rats typically repeat a lick in clusters that are separated by pauses. It has been demonstrated that rats licks more within each cluster when consuming higher concentration sucrose solutions (Spector, Klumpp, & Kaplan, 1998). Thus, the number of licks within each cluster is a reliable measure of palatability. As may be predicted by the hypothesis that effort required to obtain food increases the palatability of the food, rats showed a reliably larger number of licks per cluster when consuming the same sucrose solution following 50 lever presses (high effort condition) than following 10 lever presses (low effort condition). In a separate condition, the researchers found an analogous, but less strong, effect of a delayed delivery of the sucrose solution, with the number of licks per cluster increasing as a function of the delay. It appears that the delay itself presented an obstacle to the goal of sucrose consumption.

In explaining these animal findings, a perceptual contrast mechanism has been proposed. That is, the same reward may appear more palatable when presented while the animal undergoes a negative internal state (Zentall, 2010). The biosocial model agrees with this point of view in general outline. In addition, however, it offers a specific affective mechanism by which this contrast effect may come about. When a delay occurs in the delivery of a reward, there arise a negative reward prediction error, which disrupts the approach behavior. The resulting behavioral conflict is negatively arousing, and, as a result, the animal seeks additional incentives in the reward being pursued.

3.6 Learning from Addiction

Building on the analysis of animal dissonance, the biosocial model suggests a close link between dissonance and addiction. For example, consider chronic gamblers in Vegas or any of numerous pachinko parlors in Tokyo. Among these individuals, the prospect of winning exists hand in hand with repeated losses. The gamblers are said to be addicted to gambles because they cannot stop gambling despite the fact that they repeatedly suffer monetary losses. However, the biosocial model suggests that the excessive attraction they exhibit toward the act of gambling may have an important origin in this very fact. That is, when the prospect of winning is combined with repeated failures to win, the combination produces acute behavioral conflict. As argued earlier, this negatively arousing state initiates an active search for positive incentives linked to an option to select and pursue. In all likelihood, then, positive affect will follow the conflict between the prospect of winning and
the reality of repeated loss, and this affect will be conditioned to all cues associated with the goal of winning. These cues include slot machines, noise they generate, as well as neon signs and ambient lighting that often exist in Vegas and Atlantic City. The same mechanism may account for other forms of addiction, such as smoking, eating sweets, and reckless driving, to name just a few. For many smokers, for example, anticipated sensory pleasure stemming from a puff of a cigarette is strongly conditioned to otherwise neutral cues, such as the smell of coffee.

Flagel and colleagues have shown that conditioning of strongly appetitive motivational state to cues associated with a valued goal state (e.g., food) can easily be established in rodents (Flagel et al., 2010). Analogous conditioning has been demonstrated with pigeons (Zentall & Laude, 2013). In terms of neurophysiological mechanisms, these conditionings among non-human animals and the gambling and other forms of addiction in humans are analogous (Tomie, 1996). In fact, as shown by Flagel et al. (2010), among some sizable subgroup of rodents, the conditioning is so pronounced that they end up licking the response lever even when no food is forthcoming. Robinson and colleagues have called these animals “sign trackers,” because they are attracted to cues predicting rewards rather than to the rewards themselves. The “sign trackers” are distinguished from “goal trackers,” who are attracted only to rewards when they are presented (Robinson & Berridge, 2001). More recent work show that uncertainty associated with reward delivery (which likely produces a greater response conflict) increases sign-tracking behaviors in rodents (Anselme et al., 2013). The animals appear to be addicted in the sense of exhibiting excessive attraction to goal-related cues (the lever) under conditions of uncertainty or even extremely low probability of achieving the goal itself. As may be predicted, the sign-tracking rodents are much more susceptible to addiction than their goal tracking counterparts. Consistent with the biosocial model, it has been argued that this “sign-tracking” behavior is mediated by dopaminergic reward-processing pathways (Robinson & Berridge, 2001).<sup>5</sup>

<sup>5</sup> Human evidence on this point is limited. Moreover, where evidence exists, it is inconsistent. Whereas some studies show increased reward processing for problem gamblers relative to healthy controls (Oberg, Christie, & Tata, 2011), others show blunted reward processing among such individuals (Balodis et al., 2012). On the basis of the biosocial model, we might speculate that the results can diverge depending on whether increased motivational salience is conditioned to a gamble itself (which would increase reward processing within the gamble) or to cues associated with the gamble such as slot machines and ambient lighting (which could diminish reward processing within the gamble).
There is a lot to learn from the addiction research discussed here. This literature suggests that addiction and affective decision making share one important component in common, namely, the search for positive incentives to counter negative affect generated by delayed reward, denied views from the mountaintop, and perhaps normatively prohibited pleasures of smoking or drinking. Admittedly, addiction to illicit drugs is typically both more acute and intense than dissonance. Notably, certain illicit drugs, most notably cocaine, are likely to produce both pleasure and aversion (Ettenberg, 2004). Paradoxically, this aversion may produce behavioral conflict with the appetitive tendency, which may augment the pleasure associated with it.

Aside from obvious cases of addiction, affective conditioning and motivational cravings that result from such conditioning is quite commonplace across many domains of life and may be applicable to more mundane practices of social life. Humans are said to be symbolic animals (Lakoff & Johnson, 2003), meaning that we are deeply attached to, and strongly motivated by, symbols of significance. It may not be too far-fetched to suggest that we track symbols of our society and culture as eagerly and vigorously as Robinson’s sign-tracking rodents lick the levers of their cages.

3.7 Ego-Depletion

Over the last two decades, many psychologists have been fascinated by self-control and the failure thereof (Heckhausen & Schulz, 1995; Kross, Ayduk, & Mischel, 2005; Mischel, Shoda, & Rodriguez, 1989). One influential analysis utilizes a metaphor of self-regulation as being based on a limited resource pool—the pool of capacities that can be translated into “willpower” (Baumeister et al., 2007). As the primary proponents of the theory argue, “the exertion of self-control ... depend[s] on a limited resource.” Thus, “just as a muscle gets tired from exertion, acts of self-control cause short-term impairments in subsequent self-control, even on unrelated tasks (Baumeister et al., 2007, p. 351).” This phenomenon has been called “ego-depletion.” The theoretical framework proposed to account for the effect is intuitive and indeed has proven to be highly generative of empirical work over the last two decades. A large number of studies show that once working on a first, relatively taxing task, individuals show reduced performance in subsequent tasks even when the latter are unrelated to the first task. It appears then that the self-regulatory resource was depleted after the first task, with little left for the subsequent tasks. However, as argued by critics of the theory, there is no solid evidence for the presence of such a
limited resource pool that is thought to be analogous to a muscle (Inzlicht & Schmeichel, 2012). Moreover, evidence is emerging that the depletion effect depends much on how individuals construe the nature of the situation in which they work (Job, Dweck, & Walton, 2010), calling into question the existence, let alone the involvement, of anything like a limited resource pool for self-control.

In an important effort toward clarifying mechanisms underlying the depletion effect, Inzlicht and Schmeichel (2012) presented a potential motivational mechanism for it. As these researchers argue, “initial acts of self-control shift people's motivation away from further restraint and toward gratification (p. 453).” The reason for this motivational shift is thought to involve specific construals on what is expected on them as experimental participants. According to Inzlicht and Schmeichel (2012), “when participants ... work hard, ... [they] may feel that they have done their part for the experiment, that they have fully met their commitment to the study, and are in fact ‘owed’ a break.” This “self-indulgence” reduces the individuals’ motivation to work on tasks administered in the study, thereby shifting their attention to gratification of potentially available pleasures. It is possible that higher-order inferences regarding the nature of obligations as an experimental subject may well be involved at least under certain circumstances. Moreover, at least in such cases, “self-indulgence” may be an accurate description of how subjects feel vis-à-vis the tasks required in the study.

It is not clear, however, whether specific construals such as obligations to the experimenter or owing a break are always involved in all of the numerous studies that show the depletion effect. The biosocial model offers a more parsimonious account of how the motivational shift can happen after working on a self-control task. When individuals work on a self-control task, the task by definition is boring and/or demanding. There is a temptation of quitting the task while completing it. This behavioral conflict is likely to automatically produce (by virtue of the link from ACC to reward-processing regions involving vSTR or Nacc) a search for positive incentives. Indeed, depleted individuals are strongly attracted to seemingly irrelevant cues for pleasure (Schmeichel, Harmon-Jones, & Harmon-Jones, 2010). The search for positive incentives may well distract the individuals from the focal task, especially when the task itself is not attractive. Under such conditions, it will shift the individuals’ motivation away from the task. Instead, they will be strongly attracted toward potentially pleasurable activities (e.g., eating chocolate-chip cookies or impulsive buying) when an opportunity to engage in such activities is offered (Vohs & Faber, 2007). In other words,
the mechanism of incentive search may be responsible for the shift of motivation toward gratification. That is to say, whereas Inzlicht and Schmeichel (2012) assume that motivational shift (which is caused by certain cognitive appraisals [e.g., “I did enough for this study”, “I deserve a break”]) lead to attention to extraneous incentives, the biosocial model hypothesizes that behavioral conflict inherent in demanding/boring tasks sensitizes the actor to extraneous incentives, which in turn diminishes the motivation to work on the task at hand.

So far, the ego-depletion literature is isolated from the volume of research on effort justification. At first glance, the two phenomena do not hang together: In the effort justification effect, engagement in a boring or demanding task leads to increased attraction of the task, whereas in the depletion effect, engagement in a seemingly identical task leads to decreased attraction of the task. Given the limited resource account for the depletion effect, effort justification is an oxymoron: Why is it that people who are depleted after working on a boring task become more attracted to it? From the dissonance theory perspective, the depletion effect is a mystery: Why is it that people who feel convinced of the increased value of a task after expending effort on it end up faltering in it?

The biosocial model provides a simple explanation for both phenomena. They tap on the same fundamental affective and motivational dynamic and, yet, apparent effects are diametrically opposite because very different types of incentives are made salient in the respective situations. When positive incentives are inherent in the task at hand as when one expends effort to reach a mountaintop or when one works on a seemingly boring study for science, potently positive incentives are attached to the task (e.g., a view from the mountaintop or progress in science). Under such conditions, the search for positive incentives initiated by behavioral conflict inherent in the boring task increases the attraction of the task (effort justification). However, if no obvious positive incentives are available in the task and/or if clearly appealing positive incentives are available outside of the task, the search for positive incentives initiated by the behavioral conflict will be directed at the irrelevant incentives, leading to a loss of motivation to work on the initial task (depletion effect).

The same analysis provides an alternative interpretation for the result in the $20 condition of the classic induced-compliance study by Festinger and Carlsmith (1959). Recall that, in this study, subjects were asked to make a counter-attitudinal statement for either a small or large amount of incentive (i.e., $1 or $20). In accordance with cognitive dissonance theory, the researchers argued that the belief that one was making a counter-attitudinal
statement with insufficient justification (i.e., the $1, rather than $20 condition) was dissonance arousing and, as a consequence, rationalization (as revealed in an attitude change in the direction of the statement) occurred in the $1 condition, but not in the $20 condition. This analysis rests on the assumption that the knowledge that one is getting $20 for the act provides a cognitive justification for it. It is compatibility of cognitive beliefs or the absence thereof that counts.

The biosocial model provides an alternative perspective. That is, making a counter-attitudinal statement produces negatively arousing behavioral conflict, which leads to a search for positive incentives. This search results in identification of something positive in events or objects involved in the statement, unless there are obvious extraneous incentives (such as the $20 offer). A key test of this reasoning would be to present a large extraneous incentive that is unrelated to the act itself. Whereas the traditional dissonance explanation would predict strong rationalization (because the incentive cannot be used to justify the act), the biosocial model would predict little or no rationalization (because this incentive terminates the search for positive incentives). For example, subjects in the $1 condition may be shown to engage in impulsive buying like Vohs and Faber’s (2007) ego-depleted subjects did. Under such conditions, however, no rationalization of a conflict-producing behavior (i.e., the traditional dissonance effect) should occur. To the best of our knowledge, such an experiment has yet to be done.

An important strength of the biosocial model analysis of the depletion effect is its parsimony. There is no need to refer to higher-order construals of, say, obligations as an experimental subject or a desire for self-indulgence. Instead, the single, and simple, principle of behavioral conflict initiating an active search for positive incentives accounts for the assortment of relevant experimental results. Moreover, the biosocial model allows us to recognize that the depletion effect and the effort justification effect are very similar, tapping into the same fundamental affective and motivational dynamics. It further specifies the conditions in which one or the other effect is more likely to ensue. By so doing, it enables us to understand how the depletion effect may be related to other phenomena that might initially appear distinct, including other dissonance effects and addiction. Above and beyond these benefits, the biosocial model is based on known brain mechanisms. Importantly, it does not require any extraneous conceptual baggage such as a limited resource pool for self-control.

Altogether, the biosocial model goes beyond the Inzlicht and Schmeichel (2012) and specifies the affective and motivational brain
mechanisms for the motivational shift that is evident in depletion effects. One important future direction could be to test the depletion effect among rodents and birds. The biosocial model would be refuted in favor of more cognitive accounts if these nonhuman animals failed to show similar effects.

4. TOP-DOWN REGULATION

So far, we have argued that dissonance arousal and reduction are affective and motivational in nature. The core of dissonance effects (as well as other related effects such as ego-depletion and addiction) involves a fundamental link from detection of behavioral conflict to a search for positive incentives. Cognition is secondary and only optionally involved. This, of course, does not mean that cognitions are irrelevant. They are likely to come into play in numerous different ways. They may even prove to be powerful in both impact and consequence. First and foremost, cognitions such as beliefs and values can be a source of significant behavioral conflicts, as when one’s value of egalitarianism causes dissonance if the person is led to act in racially or sexually prejudiced fashion (Amodio et al., 2004). Second, the biosocial model also holds that higher-order cognitions sometimes regulate the system of conflict detection and, thus, modulate the magnitude of dissonance. It is this second possibility to which we now turn.

4.1 Brain Mechanisms

Imagine you are making a decision that has immediate consequences on your life. For example, you might be choosing between two cars because you need one quite soon to start commuting to your new job. You may be much more careful than usual this time, because you have to take into account various parameters including distance to the job, road conditions, and general professional expectations at the job. These cognitions will prompt you to “tighten” the belt of your thinking and planning. What this means, in neuroscience terminology, would be that you upregulate the sensitivity of your conflict detection system so as to minimize any errors in calculation. More specifically, the person will use her cognitive understanding of the situation to regulate her own psychological system of conflict detection located in dACC. One regulatory region that is neurally connected with dACC is the dorsolateral prefrontal cortex (dlPFC) (Paus & Alamancos, 2001; Richeson et al., 2003). Thus, when higher-order
cognitions about one’s need and desire to act are activated, these cognitions may engage dlPFC, which may in turn activate dACC. This regulatory pathway is called “top-down” because it is regulated by higher-order goals. These goals influence a neocortical regulatory region of the brain (dlPFC), which exercises its control over a cortical region (dACC) and, by so doing, regulates certain subcortical reward-processing mechanisms (vSTR/Nacc).6

Existing evidence suggests that such a top-down regulation of dACC through dlPFC may occur especially when active, appetitive actions are required by the current goals. Moreover, in this top-down regulation of action, it is the left rather than right dlPFC that plays an important role. Initial evidence for this assumption comes from neurophysiological research on depression. Typically, depression is characterized by the absence of desires for appetitive tendencies of approach. Depression may then be associated with an impairment of the ability to regulate one’s appetitive behaviors. Davidson and colleagues have shown that depression is often associated with right-lateralized prefrontal activation or decreased activation of the left prefrontal cortex that includes dlPFC (Davidson et al., 2002). One interpretation of this well-documented association is that depression involves an impairment of the left dlPFC function to regulate appetitive, approach-oriented behaviors. In support of this analysis, several studies have shown that repeated magnetic stimulations of the left dlPFC often leads to a reduction of depressive symptoms that are resistant to antidepressive drugs (George et al., 1997). Paus and Alamancos (2001) showed that repetitive transcranial magnetic stimulation of the left dlPFC resulted in increased cerebral blood flow in the ACC area and, on the basis of this finding, reasoned that the stimulation of the left dlPFC may compensate the impaired ability of dACC. According to Paus and Alamancos (2001), one important function of the left dlPFC is to sensitize dACC, and the increased dACC function is thought to be necessary to carry out effective appetitive behaviors.

Altogether, we may hypothesize that when individuals are prepared to execute certain actions, their left dlPFC is engaged to exercise top-down control over dACC and to sensitize the individuals to behavioral conflicts. In the present context, this means that when individuals are prepared to

---

6 Bear in mind that the higher-order goals themselves may be recruited by certain subcortical processing, such as the one implicated when one is identifying which goals may be most rewarding or dangerous. The subcortical processing in this case, however, is providing input to the higher-order regulatory processes.
execute certain actions, their conflict detection system may be sensitized so that the individuals are more likely to take note of and recognize existing conflicts in their decisions and behaviors, resulting in a greater magnitude of any dissonance that is induced (see the upper left corner of Figure 1).

4.2 Action Orientation

Preparedness for action has been referred to as action (vs. state) orientation (Kuhl, 1992) or, equivalently, implemental (vs. deliberate) mindset (Gollwitzer, Heckhausen, & Steller, 1990). Kuhl (1992) conceptualized the distinction as a relatively stable dispositional variable. Action-oriented people are thought to implement and execute their behavioral intentions more effectively. They tend to be more impulsive and behaviorally active. In contrast, state-oriented people exhibit a myriad of effects (e.g., rumination, deliberation, and anxiety) that impair their ability to implement actions. They are characterized by inability or unwillingness to pursue action with any vigor and, instead, tend to be more deliberative and even contemplative. The same dimension can vary as a function of situational demands and temporary goals (Gollwitzer et al., 1990; Kuhl, 1981). In fact, numerous studies have tested the implications of temporary activation of action versus state orientations for various behavioral outcomes, including dissonance and dissonance reduction (Harmon-Jones et al., 2009).

According to the biosocial model (Figure 1), when action orientation is evoked, and thus the left dlPFC is engaged, it will upregulate dACC, thereby sensitizing the decision maker to existing decision conflicts. It may be anticipated that action orientation should amplify dissonance and, thus, dissonance-motivated attitude change. Beckmann and Kuhl (1984) measured action versus state orientation with a self-report scale and examined whether this dispositional action orientation would moderate a choice justification effect (Beckmann & Kuhl, 1984). Participants were asked to rate the attractiveness of 16 different apartments both before and after making a hypothetical renting decision. As in numerous free-choice dissonance experiments, the preference for a chosen apartment increased after the choice. However, this choice justification effect was reliably larger for those high in dispositional action orientation.

Analogous effects have been obtained when action (vs. state) orientations are induced experimentally via priming procedures. In one such experiment, Harmon-Jones and Harmon-Jones (2002) had participants rate various aerobics exercises and then either make an easy choice (between a highly

Author's personal copy

104 Shinobu Kitayama and Steven Tompson
rated exercise and a lowly rated exercise) or a difficult choice (between two similarly rated exercises). Following the choice, participants were instructed to either list seven things that they do in a typical day (control condition) or seven things that they could do to improve their performance of the exercise (action-oriented condition). As predicted, the researchers found that participants increased their preference for the chosen option and decreased their preference for the rejected option following a difficult, but not an easy, choice. Importantly, this effect was larger for the individuals primed with an action orientation than for those in the control group. Note that, in this experiment, action orientation was induced after the choice. Hence, this orientation increased the choice justification effect after, rather than during, the choice. The biosocial model accommodates a finding like this by hypothesizing that the decision mechanism that leads to choice justification can be recursively engaged (Figure 1). We will return to this point later in the current article.

The biosocial model holds that when action orientation is induced, it will engage the left dlPFC, which sensitizes dACC, thereby increasing sensitivity to potential conflicts in the decision. Evidence for this assumption comes from a series of studies by Harmon-Jones et al. (2008). In one study, the researchers found that priming action orientation led to both increased left frontal EEG asymmetry and increased choice-induced dissonance (Harmon-Jones & Harmon-Jones, 2002). Recall that the activation of the left dlPFC is reflected in the left frontal EEG asymmetry (Sutton & Davidson, 1997). Hence, this evidence strongly suggests that the left dlPFC activity may significantly modulate the dissonance effect.

Critically, the researchers directly manipulated the left frontal EEG asymmetry using neurofeedback training (Allen, Harmon-Jones, & Cavender, 2001) and tested whether this experimentally manipulated left frontal EEG asymmetry would be related to choice-induced dissonance. The study consisted of three sessions (spread across 3 days) where the participant was given direct feedback in the form of auditory signals to indicate whether their relative left frontal EEG asymmetry was moving in the desired direction (Harmon-Jones et al., 2008). Half of the participants were given training to increase their left frontal EEG asymmetry, whereas the other half were given training to decrease their left frontal EEG asymmetry. Participants initially received 2 days of neurofeedback training. On the third day, they rated their preference for psychology experiments, chose between two equally preferred options, received their third session of neurofeedback training, and then rated the options again. The results showed a powerful
effect of neurofeedback. As shown in Figure 7, individuals who had been trained to increase their left frontal cortical activation showed a clear choice justification effect. In contrast, this effect disappeared for those who had been trained to decrease their left frontal cortical activation. The finding suggests that the left dlPFC causes choice-induced dissonance to increase. This is consistent with the hypothesis that the activation of the left dlPFC activates dACC, which sensitizes the decision maker to potential decision conflicts.

4.3 Arousal or Reduction?
Harmon-Jones and colleagues (2008, 2009) have explained their findings in terms of their action-based model of dissonance, which postulates that decision conflict impedes an effective execution of an action. Thus, when action orientation is enhanced, the conflict becomes especially aversive, thereby strongly motivating the decision maker to resolve the conflict. In other words, the reason why the decision maker would try to resolve the decision conflict is to achieve the goal of executing effective actions. It would follow, then, that in the absence of any immediate need or desire for action.

Figure 7 When the left frontal EEG asymmetry is increased via neurofeedback, a clear justification effect is observed (the right panel). That is, preferences for chosen and rejected items spread after choice, but this effect is eliminated when the left frontal EEG asymmetry is decreased via neurofeedback (the left panel). Reprinted from Harmon-Jones et al. (2008) with permission from APA.
execution, the dissonance effect should dissipate or at least should become weaker. Action orientation is thus held to modulate dissonance reduction.

The biosocial model holds that when action orientation is induced, it sensitizes the conflict detection system as implemented in dACC. Thus, when oriented toward action, individuals will be especially sensitized to available decision conflicts. Once the conflicts are registered, they cause negative arousal, which motivates the decision maker to find positive features in one of the available decision options. When action orientation is weak and/or a contrasting state of state orientation is dominant, the sensitivity of the dACC conflict detection system is lowered. Under such conditions, the decision maker may not take note of existing decision conflicts. Hence, there will be no dissonance effect either. These implications of the biosocial model are consistent with the action-based hypothesis, with a few points of possible disagreement.

To begin, the action-based model implies that it is the effort to reduce dissonance that is modulated by action versus state orientation. Conflicts inherent in a choice supposedly produce the dissonance regardless of whether action orientation has been induced or not. Thus, state-oriented people do not feel motivated to reduce the dissonance that has been induced by the conflicts. In contrast, the biosocial model implies that action or state orientation influences the dACC sensitivity and, thus, it modulates the magnitude of dissonance itself. Indeed, the biosocial model suggests that as long as strong behavioral conflicts are detected, they are likely to activate negative arousal, thereby motivating the decision maker to look for positive features unique to one of decision options. This prediction may be tested with neuroimaging methods, with the activation of dACC and aINS as an index of dissonance arousal (see Figure 3).

Another potential point of disagreement concerns the relative weight the two analyses place on action orientation. The biosocial model acknowledges that action orientation is one factor that increases the sensitivity of dACC. However, it also suggests that there are many other factors that modulate the sensitivity of dACC. Some regulate the sensitivity of dACC through top-down regulatory pathways, including the one mediated by the left dIPFC. These factors include action orientation. But there may be other top-down factors that regulate the magnitude of dissonance. For example, social situations vary in the latitude of normatively permissible behaviors. Thus, some settings, such as funerals and libraries, are much narrower in the latitude of tolerance and, thus, more tight than others, such as public parks and rock concerts (Gelfand et al., 2011). More careful top-down monitoring of one’s
behavior may be called for in tight social situations. We may then anticipate that individuals should experience greater dissonance in tight rather than loose situations, net of action orientation. At this point, we are unaware of any data speaking to this prediction. Nor do we know whether the effect of situational tightness might also be mediated by the left dIPFC. Future work should address this and other factors that regulate the sensitivity of dACC through top-down pathways.

The biosocial model also assumes that in addition to the top-down factors discussed in this section, there likely exist many significant factors that regulate the sensitivity of dACC bottom-up, through different neural pathways (see the lower left corner of Figure 1). These pathways are likely to be just as important as the top-down pathways, including the one mobilized by action orientation. We now turn to the bottom-up regulatory pathways of dissonance.

5. BOTTOM-UP REGULATION

5.1 Brain Mechanisms

The dACC functions may be regulated, bottom-up, by cues signaling either safety or threat. When individuals are reassured of their safety, they can relax, because there is no need to be vigilant for any potential contingencies in the environment. Thus, safety cues downregulate dACC, thereby reducing the magnitude of behavioral conflict (i.e., dissonance). In contrast, when the individuals face difficulties, risks, and potential dangers, they will be threatened. Such threat cues will mobilize neural mechanisms that enable them to be vigilant for any potential environmental contingencies that pose a threat. Threat cues upregulate dACC, thereby increasing the magnitude of behavioral conflict (i.e., dissonance).

The regulatory path illustrated above is called “bottom-up” because the cues of safety or threat are detected in subcortical regions (as shown below) and take control over a cortical region (dACC). It should be born in mind that these cues themselves may often require substantial neocortical processing, such as when one is exposed to religious symbols or one’s sense of the self is affirmed as we shall see below. However, it still remains the case that the primary source of regulation is located in the subcortical regions; thus, the regulatory pathway itself is directed upward from these regions.

Previous work suggests certain plausible brain mechanisms for the bottom-up regulation of the dACC function. In particular, safety cues may down-regulate dACC through subcortical dopaminergic (DA) pathways. These
pathways originate in specific midbrain areas (the substantia nigra and the ventral tegmental area). DA neurons of these regions are strongly activated by positive incentive cues. This DA burst is likely to spread to the striatal reward-processing region as noted earlier (see Section 3.2). More importantly in the present context, it can also spread to dACC. In this case, the DA burst in midbrain may be expected to decrease the dACC functions, because the connection between the DA pathway and the dACC is largely inhibitory (Bromberg-Martin & Hikosaka, 2009; Walsh & Anderson, 2012). This implies that positive incentive cues, especially those that are likely to result in relatively tonic increase of a positive state, may result in inhibition of the dACC sensitivity and, thus, to a reduced ability to recognize available conflicts in one’s behavior and decisions. Although this analysis is speculative and must be examined further in future work, it provides a reasonable neural mechanism that depresses the ACC sensitivity as a function of relatively tonic, positive incentive cues including those signaling safety.

Often times, the world is capricious, full of unexpected happenstances. And, of course, people may also be frightened by the prospect of their own death. According to terror management theory (Greenberg, Solomon, & Pyszczynski, 1997), one cultural instrument humans invented to offer an assurance of eternal life and salvation is religion (Norenzayan & Shariff, 2008). If the hypothesis above is correct in that safety cues downregulate dACC function, religious ideations may be sufficient to cause a tonic deactivation of dACC, especially for those with religious faith. A series of ERP studies reported by Inzlicht and colleagues have provided initial support for this prediction (Inzlicht, McGregor, Hirsh, & Nash, 2009; Inzlicht & Tullett, 2010).

In this work, the researchers examined ERPs of Canadian participants, who worked on a standard color-naming Stroop task. The participants were shown a series of color words presented in a color that either matched or mismatched the semantic meaning of the word. They indicated the font color by pressing a colored response key while ignoring the semantic word meaning. When individuals make an error in a cognitive conflict task like this, their ERPs, especially those in the midline frontal area, show a notable negative-going spike nearly simultaneously with the initiation of the response. As noted earlier, this ERP component is called ERN (Gehring & Willoughby, 2004; Holroyd & Coles, 2002). ERN is thought to originate in dACC (Bush et al., 2000). It signals two types of conflict or error. First, it can signal response conflict between a response that has just been initiated and a correct response that is based on cognitive processing
of the word meaning (Botvinick et al., 2004; Yeung, 2004). Second, it can also signal a negative reward prediction error (Walsh & Anderson, 2012), which occurs when the actual outcome (i.e., error) is worse than expected (i.e., correct response). The negative reward prediction error interrupts the behavior and, thus, can be seen as a go- versus no-go-type behavioral conflict. Both of these computations have been localized to dACC or nearby regions. Thus, when the dACC sensitivity is reduced, the ERN magnitude should be decreased.

Data provided support for the above prediction. In their first set of experiments, Inzlicht et al. (2009) measured “religious zeal” by using a self-report scale (e.g., “I aspire to live and act according to my religious beliefs,” “My religious beliefs are grounded in objective truth”) and found that the ERN magnitude systematically decreased as a function of religious zeal. In their subsequent studies, the researchers primed positive religious ideations directly and tested whether individuals would manifest less ERN (Inzlicht & Tullett, 2010). As predicted, ERN was weaker in the religious priming condition for religious participants but not for atheist participants (who would not be expected to view religion as a safety cue).

Conversely, when safety cues are removed, when they are psychologically absent, or when the level of anxiety is increased, dACC function should be increased. There is increasing evidence that the ERN magnitude increases as a function of trait anxiety (Hajcak, 2012). Moreover, numerous studies have shown that avoidance-related motivational states influence ERN (Gehring & Willoughby, 2004; Holroyd & Coles, 2002). For example, when motivation is increased by monetary incentives or by the presence of another person evaluating the participant’s performance, the ERN magnitude is increased (Boksem, Tops, Wester, Meijman, & Lorist, 2006; Hajcak, Moser, Holroyd, & Simons, 2006).

Pertinent neural pathways by which anxiety or certain other threats upregulate the dACC function are largely unknown. One possible pathway implicates the amygdala. Impinging threats are likely to sensitize ACC (Öhman, 2005) via the amygdala, which is known to respond strongly to threat cues (Nader, Schafe, & Le Doux, 2000; Öhman, 2005). It is possible that the amygdala activation directly sensitizes the dACC. Much caution is justified here, because the amygdala responds to rewards as well as fears or threats, subserving motivational functions that are far more fine-grained and nuanced than previous theorizing implied (Robinson, Warlow, & Berridge, 2014).
In addition to the direct link from the amygdala to ACC, anxiety or fear detected at the amygdala may influence ACC through an indirect route. As noted above, dACC is linked through inhibitory connections to the dopaminergic (DA) reward pathway, which originates in the midbrain. The firing rates of the midbrain DA neurons are likely to decrease upon the detection of a threat cue (Bromberg-Martin & Hikosaka, 2009; Frank & Claus, 2006). The decreased activation of the DA pathway, in turn, may be expected to disinhibit the ACC because, as noted, the connection of the DA pathway to the ACC tends to be inhibitory (Walsh & Anderson, 2012). In this way, threat cues may increase dACC sensitivity.

5.2 Generalized Other and dACC Sensitivity: A Cultural Variation

Some signals of safety such as religion may be universally available across cultures (Norenzayan & Gervais, 2012). Likewise, some signals of threat such as loss of monetary reward may be universal. However, many cues of both safety and threat may also be inherently entrenched in divergent cultural practices and meanings. Effects of such cues may be highly variable across cultures. One such cue may involve images of what George Mead called the “generalized other” (Mead, 1934).

It has been proposed that cultures vary systematically in terms of the model of the self that is shared and authenticated therein (Kitayama & Uskul, 2011; Markus & Kitayama, 1991). In Eastern cultures (especially East Asian cultures), the self is assumed to be interdependent. According to this model of the self, one’s behavior is guided and organized by others’ expectations and obligations to them. For these individuals, evaluations by others are so important that when interacting with others, they are very likely to worry about these evaluations. In fact, Asians typically attend closely to their potential shortcomings and negative features (Kitayama, Markus, Matsumoto, & Norasakkunkit, 1997) and, thus, tend to be more pessimistic (Chang & Asakawa, 2003) and higher in social anxiety (Okazaki, Liu, Longworth, & Minn, 2002). This pessimism enables the individuals to maintain positive evaluations in the eyes of the others (Heine, Kitayama, & Lehman, 2001). Once socialized in this interdependent cultural system, individuals may associate certain negative emotions such as worry, apprehension, and anxiety with an image of a “generalized other” (Mead, 1934). Accordingly, face cues may acquire the potential to upregulate the dACC functions.
In contrast, in Western cultures, especially European American cultures, the self is assumed to be independent. According to the independent model of the self, one’s behavior is guided and organized by his or her internal attributes such as desires, attitudes, and preferences. These individuals tend to focus on positive aspects of themselves (Kitayama et al., 1997) and are thus optimistic (Taylor & Brown, 1988). As may be expected, European Americans tend to be relatively low in social anxiety (Okazaki, 2000) while showing markedly higher levels of general trust as compared to Asians (Yamagishi, Cook, & Watabe, 1998). Once socialized in this cultural context, individuals may acquire contrastingly positive associations with images of the generalized other. These images may then be accorded the potential to serve as safety cues and inhibit the dACC functions.

In a recent experiment, Park and Kitayama (2014) had both European American and Asian participants perform a speeded flanker task responding to the direction of a center arrow flanked by either congruent (same direction) or incongruent (opposite direction) arrows (Park & Kitayama, 2014). Right before the flanker was presented, a realistic face image that was both racially and emotionally neutral was presented very briefly for the average duration of 90 ms. As expected, the face priming changed the magnitude of the ERN. For Asians, the ERN was significantly larger in the face-priming condition as compared to control conditions where either an image of a house or a scrambled face was presented as a prime. This finding is consistent with the hypothesis that face priming sensitizes dACC. In contrast, face priming tended to decrease the magnitude of ERN for European Americans, suggesting that face priming desensitizes dACC. The pertinent waveforms are shown in Figure 8A and B for Asians and European Americans, respectively.

Note that in this experiment, the priming stimuli were presented for less than one-tenth of a second. Thus, although the primes were visible for the most part, they were very brief. Thus, it is unlikely that participants clearly registered them or exercised any effort to regulate their actions because of them (Cunningham et al., 2004). The pattern of data thus suggests that culture establishes quite divergent emotional conditionings to the generic face image, consistent with the hypothesis that the positive versus negative associations the face cues have for the two cultural groups of participants are due to the underlying dimension of interdependence (as opposed to independence). If this hypothesis is correct, the cultural difference may be explained by individual differences on this dimension. Figure 8C shows the magnitude of the face-priming effect on ERN as a function...
As can be seen, as one’s level of interdependence becomes higher, the face-priming effect on ERN becomes more positive. The data are consistent with the hypothesis that Asians show a positive face-priming effect (with the face increasing the ERN magnitude) because they are relatively higher in interdependence, whereas European Americans show a negative face-priming effect (with the face decreasing the ERN magnitude) because they are relatively independent.

Another paradigm suitable in testing error processing involves a gambling task. Hitokoto, Glazer, and Kitayama (2014) presented their participants with two boxes on the computer screen and asked them to choose one to obtain monetary points. Right after the choice, feedback was delivered regarding whether they had gained or lost certain points. This gamble was repeated a number of times. When the gamble was presented on each trial, a priming stimulus (either face or scrambled face) was presented very briefly, for 90 ms, between the two boxes. The goal was to determine whether the face priming would also modulate ERP error signals. In a gambling task like this, approximately 270 ms after the feedback, there emerges a

Figure 8 ERN magnitude as a function of culture and face priming. (A) Wave forms for Asians. The ERN peak is higher in the face-priming condition than in the control conditions. (B) Wave forms for European Americans. The ERN peak is lower in the face-priming condition than in the control condition. (C) The face-priming effect is predicted by interdependent (vs. independent) self-construal. Reprinted from Park and Kitayama (2014) with permission from Oxford University Press.
negative deflection of ERP around the frontal central scalp location when
the outcome is negative and, thus, worse than expected. This signal is called
feedback-related negativity (FRN) and is analogous to ERN in that it is an
index of the detection of a negative reward prediction error (Gehring, 2002;
Gehring & Willoughby, 2004; Miltner, Braun, & Coles, 1997).

Hitokoto et al. tested whether the magnitude of FRN would be
moderated by face priming, with an expectation that the FRN magnitude
would increase as a function of face priming for Asians, but it should
decrease as a function of face priming for European Americans. This, in fact,
was the case with one caveat. In this study, two different groups of Asians
were tested, Asian Americans (i.e., individuals with Asian heritage who
had been born and brought up in the United States) and Asian sojourners
(i.e., those with Asian heritage who had been born and brought up in Asia
and who came to the United States relatively recently). As shown in
Figure 9, the face-priming effect on FRN was positive for Asian Americans,
but negative for European Americans, as predicted. Interestingly, the pattern
for Asian sojourners was more similar to the one for European Americans

Figure 9 Magnitude of FRN (feedback contingent negativity on loss vs. win trials) in the
face priming (as compared to scramble face control) condition in a gambling paradigm.
For Asian Americans, face priming significantly increased the magnitude of FRN (the
magnitude of negativity on loss vs. win trials), whereas for European Americans face
priming tended to decrease it. This effect achieved statistical significance once trials
with bursts of alpha wave were excluded. Unexpectedly, Asian sojourners showed a pat-
ttern that is no different from the European American pattern. Taken from Hitokoto et al.
(2014).
than for Asian Americans. This finding must first be replicated before being taken seriously. Tentatively, we may speculate that Asians who chose to come to the United States for college (and then chose to participate in a neuroimaging study under certain recruitment procedures and conditions) might be self-selected to be more independent than may be typical among Asians.7

5.3 Culture, Face Priming, and Dissonance

The ERP studies discussed earlier by both Inzlicht and colleagues on the effect of religion priming (Inzlicht & Tullett, 2010) and by Kitayama and colleagues on the face priming (Hitokoto et al., 2014; Park & Kitayama, 2014) offer some significant implications for the biosocial model, because the ERP signals tested in these studies (ERN and FRN/FRP) are thought to originate in dACC and serve as reliable indices of the dACC sensitivity. To the extent that dissonance effects are also mediated by the ACC sensitivity, there should be comparable effects of priming on the magnitude of dissonance effects.

In a series of cross-cultural free-choice dissonance studies, Kitayama and colleagues have provided substantial evidence for this hypothesis (Imada & Kitayama, 2010; Kimel et al., 2014; Kitayama et al., 2004). In a typical free-choice dissonance study, participants are asked to rate several choice options both before and after they make a choice between two of the options. The two options happened to be options that were rated similarly during the first prechoice rating period. Numerous studies conducted in North America have found that after the choice, attraction to the chosen CD increases and attraction to the rejected CD decreases (e.g., Harmon-Jones et al., 2009). As noted earlier, Kitayama et al. (2013) provided the first neuroimaging evidence indicating that a difficult choice like this produces a conflict as indicated by dACC + aINS activation, which in turn prompts the chooser to find positive distinctive features in one of the options. Once these positive distinctive features have been identified (as indicated by the Nacc activity), the chooser will make a clear choice (see Figures 1 and 5). After the choice,

---

7 The aforementioned study by Park and Kitayama (2014) also tested Asian sojourners and found their pattern to be contrastingly different from the pattern for European Americans (Figure 8C). Moreover, cross-cultural dissonance studies available in the literature (to be discussed in Section 5.3) tested Asians in Asia, Asian Americans, and Asian sojourners and found patterns that are, for the most part, very similar to one another, although no study directly compared Asian Americans and Asian sojourners within a single study as in the Hitokoto et al. (2014) study.
the chosen CDs become more attractive because of this newly identified positive information associated with it.

Importantly, an earlier study had shown that under this standard free-choice situation, Asians do not show the choice justification effect (Heine & Lehman, 1997). The two ERP face-priming studies (Hitokoto et al., 2014; Park & Kitayama, 2014) suggest why this might be the case. In the absence of any watching faces, the dACC of Asians might not be sufficiently activated. Perhaps, due to anxiety associated with images of others, the absence of such others may serve as a potent cue for safety. Importantly, Asians should show a reliable choice-justification effect once they are exposed to a face cue while making a choice. Conversely, European Americans should show the choice-justification effect in the absence of any face cues, but they might not show it once they are exposed to face cues (signaling safety) during choice. In one study (Kitayama et al., 2004), both Asian and European American participants rated 10 popular music CDs both before and after making a choice between two CDs that had initially been rated as nearly equally liked. A poster (seemingly prepared for a conference presentation) was inconspicuously displayed on the wall in front of half of the participants (Figure 10A). The poster showed several schematic faces that appeared to be “watching” the participants. The remaining half of the participants were not exposed to this face poster.

As predicted, there was a significant face × culture interaction, which is illustrated in Figure 10B. Replicating the earlier study by Heine and Lehman (1997), Asians did not show any justification effect (i.e., increased liking for

![Figure 10](image-url)
the chosen CD and decreased liking for the rejected CD) in the no-poster condition. However, once they were exposed to the face cues, these participants showed a highly significant choice justification effect. The finding is consistent with the hypothesis that Asians associated anxiety and worry with face cues. Moreover, they appear to take the absence of face cues as signaling safety. This finding is also consistent with the biosocial model, suggesting that the face cues sensitized dACC, initiating the decision mechanism that assigns positive affect to the chosen CD. Equally importantly, European Americans showed a reliable justification effect in the absence of the face cues, replicating numerous previous studies. However, this effect became weaker in the face-priming condition. The reduced justification effect in the face (vs. control) condition has since been replicated (Imada & Kitayama, 2010; Kimel et al., 2014). The finding suggests that European Americans associate a positive incentive of safety with an image of the generalized other. The face cues therefore desensitize dACC, leading to reduced conflict detection and choice justification.

In a subsequent study, Imada and Kitayama (2010, Study 2) showed that for the face-priming effect to occur, it is critical that face representations must be activated when exposed to the prime. In this study, Asian and European Americans were tested in a free-choice dissonance paradigm. Instead of the face poster, the researchers used a letter-size sheet of paper on which three dots were printed in either a triangular formation or a reversed triangular formation (see Figure 11A). This sheet was inconspicuously placed in front of the participants. After the free-choice procedure, the experimenter took up the sheet and asked the participants to complete a picture by using the three dots. As shown in Figure 11B, some participants drew a face by using the three dots, whereas some others drew some other picture. Conceptually replicating the Kitayama et al. (2004) study, the choice justification effect was significantly greater for Asians who drew a face than those who did not. Conversely, the effect was significantly greater for European Americans who did not draw a face than those who did (Figure 11C).

In both the Kitayama et al. (2004) study and the Imada and Kitayama (2010) study, face-priming pictures were presented throughout the experimental procedure. A recent study tried to sharpen the analysis by examining the critical time window in which the face-priming effect would be maximal. According to the biosocial model, the face cues are considered to modulate dACC sensitivity, which initiates the decision mechanism that produces positive affect assigned to an option to be chosen. This process can be complete before a decision is made, although the process
may sometimes be recursively engaged. This analysis implies that the face-priming effect observed in the two studies discussed above should be more pronounced when participants are exposed to face cues while they make a choice rather than after the choice.

To examine this issue, Imada and Kitayama (2010, Study 1) used the original face-priming procedure of Kitayama et al. (2004). Unlike in the Kitayama et al. (2004) study, however, right after making a choice on one desk, participants were asked to move to another desk in the same room. The experimenter casually told them to do so because she had to clean up the first desk for another session. The face poster was placed in front of either the first desk or the second desk. Thus, the participants were exposed to the face poster either while making a choice or after the choice. As predicted by the biosocial model, Asians showed a significantly larger choice justification effect when they were exposed to the face poster while they made a choice than when they were exposed to it after the choice. In contrast, European Americans showed a completely reversed pattern, with the choice

![Figure 11](image-url) Face-priming effect in choice justification. (A) A three-dot configuration placed in front of participants. (B) Sample drawings using the three-dot configurations. Note some are faces, whereas the others have nothing to do with face. (C) For Asians, those who drew a face showed a greater justification effect, whereas for European Americans, those who drew a face showed a weaker justification effect. Reprinted from Imada and Kitayama (2010) with permission from Guilford Press.
justification effect significantly weaker when they were exposed to the poster during the choice than afterward.

The same analysis can be extended to explain cultural differences in dissonance effects as a function of the choice being made for the self versus for a close other. When individuals make a choice for themselves in a private setting, no subjective representation of the generalized other is actively available. Under such conditions, European Americans tend to show a much larger justification effect than do Asians. However, when the individuals make a choice for a close other in a public setting, images of others watching the choice (including the other person for whom the choice is being made) become highly salient. It may be expected, then, that under such conditions, Asians will show a choice justification effect more strongly than do European Americans. A series of studies by Hoshino-Browne and colleagues has confirmed these predictions (Hoshino-Browne et al., 2005). This Asian effect appears to be mediated by their self being relatively interdependent. In a recent series of studies, Kimel, Grossmann, and Kitayama (2012) show that European Americans also show this effect when primed with interdependent orientations.

Taken together, the cumulative evidence reviewed in this section shows that for Asians, face priming upregulates the ACC sensitivity and, as a consequence, increases the attractiveness of the chosen options. In contrast, for European Americans, face priming downregulates dissonance and, as a consequence, decreases the attractiveness of the chosen options. Moreover, the fMRI evidence reviewed earlier indicates that the change in attractiveness is likely to be anchored in the corresponding change in activity in the striatal reward-processing areas (Jarcho et al., 2011; Kitayama et al., 2013). It is important to bear in mind, however, these areas are involved not only in “liking,” but also in the motivational striving of “wanting”; moreover, these two aspects of reward processing are distinct, overlapping only partially (Berridge, 2012; Berridge et al., 2009). Accordingly, it is of interest to determine whether the culturally variable choice effect observed in the studies reviewed here could be extended to motivational inclinations toward chosen goals.

Another recent study investigated this possibility by having both European Americans and Koreans choose one of three tests measuring different aspects of IQ (fluid IQ, creativity, and analytic IQ) (Na & Kitayama, 2012). One-third of the participants made this choice with the face poster hung in front of them, whereas another one-third did so without any such poster. The remaining third did not make this choice, but the experimenter simply assigned one of the tests to work on. After these manipulations,
the participants moved to a separate individual booth and worked on the chosen/assigned test. Unbeknownst to the participants, they worked on an identical set of remote association questions. The researchers measured how many questions they solved within 5 min. The results are summarized in Figure 12. As can be seen, Koreans solved more questions in the face priming/choice condition than in either the no face priming/choice condition or the assignment condition. This result demonstrates that after making a choice in the presence of face priming, Asians are not only attracted to the chosen goal (as shown in the dissonance studies discussed earlier) but also motivated strongly to achieve the goal. That is, the incentive value of the chosen goal has been increased, as predicted by the hypothesis that faces serve as a threat cue for these individuals. In contrast, European Americans solved more questions in the no face priming/choice condition than in the face priming/choice condition or in the assignment condition. This shows that for European Americans the incentive-increasing effect of choice is diminished by face priming, as predicted by the hypothesis that faces serve as a safety cue for these individuals.

5.4 Self-Affirmation, Mortality Salience, and Other Bottom-Up Factors

The current analysis implies that all cues signaling safety should down-regulate dACC functions, thereby reducing the ability of the decision maker
to recognize existing decision conflicts and thus reducing dissonance effects. By extending this logic, the biosocial model provides a comprehensive framework in which to understand effects of self-affirmation (Sherman & Cohen, 2006; Steele, 1988) and mortality salience (Greenberg et al., 1997).

Self-affirmation occurs when certain positive features of the self such as moral integrity and competence are confirmed and reinforced. Cognitions required to achieve this state of affirmation may be called affirmation resources. People with sufficient affirmation resources are considered resilient, resulting in less defensive reactions when facing potentially threatening events (Sherman & Cohen, 2006). They become more resilient when they are reminded of available affirmation resources, such as high values they possess or high competence or ability of the self (Creswell et al., 2005). The biosocial model suggests that when the self is affirmed, the subcortical DA reward pathway is activated, which in turn inhibits dACC.8 Under these conditions, people will be relatively more oblivious to existing behavioral conflicts, thereby ceasing to show dissonance effects. Numerous studies have confirmed this general prediction (Heine & Lehman, 1997; Steele, Spencer, & Lynch, 1993). The affirmed individuals may also be less defensive because they are relatively oblivious to potential conflicts or errors they may commit.

One important caveat must be noted. Exactly what information serves as an affirmation resource depends on a myriad of contextual factors including personality, situation, and broader sociocultural context. For example, confirming one’s own personal values may be highly affirming to those who hold an independent sense of the self, but not to those who hold a more interdependent sense of the self. In support of this possibility, evidence indicates that choice justification is readily eliminated when one’s personal values are confirmed for European Americans (who tend to hold an independent sense of the self), but this manipulation does not seem to work for Asians (who tend to hold an interdependent sense of the self). Instead, for Asians, it is confirmation of group values (e.g., values shared by family members) that tend to have an equivalent self-affirming effect (Hoshino-Browne et al., 2005).

Moreover, every society and culture offers many other cues signaling safety. Familiarity and banality of daily life, in and by itself, may signal safety

---

8 Self-affirmation can have multiple effects. Thus, in addition to activating the midbrain DA pathway (thus, downregulating dACC), it might also motivate people to work hard on competence tasks because it enhances a general sense of self-competence. Self-affirmed individuals may then engage a top-down regulatory control to upregulate dACC (Legault, Al-Khindi, & Inzlicht, 2012). Thus, the net effect of self-affirmation on dACC sensitivity may depend on the relative strength of the two effects. This top-down factor, however, is unlikely to be relevant in many dissonance situations where no cognitive competence is at issue.
(Zajonc, 1968). Or the mere absence of potential threats (e.g., other watching others for Asians) may serve as a safety cue (as argued earlier). Likewise, certain religious rituals involving symbolic cleansing or secular practices intended to improve hygiene may also function in similar fashion. Recent work shows that an opportunity to wash one’s hands was sufficient to “wash away” a dissonance effect (Lee & Schwarz, 2010), consistent with the notion that washing hands is an effective cue signaling safety, thus down-regulating dACC, which reduces the dissonance effect.

Currently, there is little direct evidence pertaining to the neural pathways activated by self-affirming information (or for that matter any cues signaling safety) to alleviate the dissonance effects. As noted earlier, animal studies show that such positive incentive cues activate the midbrain DA neurons, which in turn downregulate dACC (Bromberg-Martin & Hikosaka, 2009). It stands to reason that individuals who are affirmed and/or relaxing in safety may show greater DA activation, which then downregulates regions including dACC. In principle, it is now in our reach to perform a critical test of the hypothesized neural pathway with available neuroimaging methods.

Another major source of dACC sensitivity is mortality salience. When mortality is made salient, people are assumed to be frightened. Although this fear is held to be unconscious insofar as it is exceedingly difficult to pin it down on the basis of subjective self-reports, it is likely to instigate the threat detection system of the brain that most likely includes the amygdala. This activation of threat may well decrease midbrain DA activity. In combination, mortality salience is likely to upregulate dACC sensitivity, thereby amplifying associated effects including error processing as revealed in ERN or FRN and dissonance reduction. As in the case of self-affirmation, this prediction by terror management theory, particularly the one pertaining to the dissonance effect, has received strong empirical support (Friedman & Arndt, 2005; Greenberg et al., 1997; Jonas, Schulz-Hardt, & Frey, 2005). At the same time, as is also true in the case of self-affirmation, the specific conditions in which this fear is instigated may depend on a number of contextual variables. For example, individualists may be more responsive to a thought of death of their personal self, but collectivists may be more responsive to a thought of death of their group (Kashima, Halloran, Yuki, & Kashima, 2004).

One benefit of an encompassing theoretical framework like the one proposed here is to highlight a common underlying mechanism that cuts across disparate, seemingly independent phenomena such as self-affirmation and mortality salience. Self-affirmation is one particular way to highlight the safety of the self, whereas mortality salience is one particular way to highlight an excruciating threat to the self. At the level of brain mechanisms involving
dACC sensitivity, these two phenomena can be seen as representing the opposite ends of a theoretical continuum defined by the degree of sensitivity of the dACC-based conflict detection system. In principle, there may be many other variables that influence this system. The biosocial model then may serve as a valuable heuristic tool for discovering new moderating variables of dissonance and related phenomena.

6. RECURSIVE LOOP

6.1 Aversive Consequences

The biosocial model implies that behavioral conflict (i.e., dissonance) is inherent in difficult decisions. Often, the conflict is aroused and resolved by the time a decision is made. In fact, this conflict itself constitutes an important part of the mechanism for decision making. However, once a decision is made, this decision may become a source of additional conflict, consistent with the notion of postdecisional dissonance. For example, the decision may go against a prevailing norm of the situation. This may alert the person to regulate her behavior more carefully, which could engage the left dlPFC to increase the sensitivity of dACC. Alternatively, the person may recognize potential aversive consequences of the decision, in which case the decision may become a cue for self-threat, mobilizing the bottom-up regulatory pathway to increase the dACC sensitivity.

Given the biosocial model, the renewed effort to resolve behavioral conflict is likely to direct the decision maker’s attention to other available positive incentives associated with the decision, insofar as this incentive information is likely to reduce the behavioral conflict and negative arousal associated with it, and enable her to resurrect her original decision. In short, this recursive loop of dissonance arousal and reduction may further amplify the justification effect. This is exactly what Cooper and Fazio (1984) demonstrated and featured in their New Look perspective on dissonance.

6.2 Misattribution

The dissonance arousal before an initial decision can be quite brief and transitory. Once aroused, it can dissipate quickly when positive incentives are identified in one of the decision options. The negative arousal is coded in terms of transitory activation of aINS. Although it powerfully regulates attention involved in the search for positive incentives, it may also be too brief to surpass the threshold for conscious awareness. For all practical purposes, the negative arousal of dissonance may be subliminal. However, once
the recursive loop of dissonance is engaged, the negative arousal may repeat itself and; moreover, the recursive loop may involve more deliberate, slow-growing psychological processes. Hence, the negative arousal of dissonance may be subjectively experienced.

Once the aINS activation reaches the conscious threshold, this activation itself may become the target of cognitive interpretations. It is possible that the decision maker correctly interprets the negative arousal as resulting from her decision conflict. However, there may be other environmental cues (e.g., a noisy room) that may reasonably explain why one might be feeling the negative arousal. Under these conditions, the decision maker may misattribute her negative arousal to the environmental cues. To the extent that such misattribution occurs, the decision maker will no longer experience any need to reduce it. The justification effect may dissipate accordingly. A large number of studies conducted by Zanna, Cooper, and many others during the 1970s provided convincing evidence for the potent effect of experimentally manipulated misattribution cues (Zanna & Cooper, 1974; see Cooper & Fazio, 1984, for a review).

6.3 Split Brain

The biosocial model holds that the conscious, interpretive aspects of dissonance arousal and experience, which are mediated by the recursive loop of Figure 1, are functionally separate and independent of the subcortical mechanism of dissonance we have identified. Intriguing evidence for this assumption comes from research on split-brain patients (Gazzaniga, 2013; Wolman, 2012). In the 1940s, to treat severe forms of epilepsy, the corpus callosum (which connects the two cortical hemispheres) was sometimes surgically dissected. The surgery helped contain the seizure to one side of the brain. In the 1950s, a small group of researchers led by Roger Sperry investigated neuropsychological characteristics of these split-brain patients. One conclusion from this body of work is that the two hemispheres have very different functions (Gazzaniga, 2013). In large part, the left hemisphere, but not the right hemisphere, is responsible for verbal, linguistic, and interpretive capacities. Among split-brain patients, subcortical mechanisms (which retain the neurological unity) are capable of automatically executing certain simple responses that are requested, whereas the left hemisphere is required to interpret what they have done and why they have done so.

One dramatic demonstration of this point comes from a patient called P.S. (Ledoux, Wilson, & Gazzaniga, 1977). In one study, for example, P.S. was flashed one picture to one visual field (so that it is projected on
the contralateral hemisphere) and asked to point to the same one from several pictures that were placed in front of him. He had no problem in carrying this out. Now imagine that two pictures are shown, with a chicken claw flashed on the right visual field (projected on the left hemisphere) and a snow scene on the left visual field (projected on the right hemisphere). Under such a condition, P.S.’s left and right hands (controlled by the contralateral hemispheres) pointed to the pictures of chicken and snow shovel, respectively. This is exactly what one would expect if each hemisphere independently responded to what it saw. Of importance for our purpose, when P.S. was asked why he did what he did, he immediately came up with a narrative that seemed reasonable, “The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed” (Gazzaniga, 2013, p. 13). The observation is consistent with a view that separate from whatever processing modules are required in the production of automatic matching responses, there exists an interpretive module, supposedly localized in the left, verbal hemisphere, which uses one’s own automatically generated responses as cues and interprets them within a coherent narrative. By so doing, split-brain patients like P.S. manage to maintain the unitary sense of the self even though much of their mind is physically split into two halves. Earlier on, we noted a demonstration of a free-choice dissonance effect among amnesic patients (Lieberman et al., 2001). It is possible that what is missing in these patients is a failure in retrospective story making (which involves episodic recollection of previous experiences), with all other dissonance machineries kept intact.

It is evident, then, that some significant portion of the higher-order system of interpretation and story making, supposedly localized in the left “verbal” hemisphere, is functionally separate and independent of the rest of the psychological mechanisms that automatically carry out various functions including dissonance and other forms of affective decision making. Thus, the specific precognitive neural network of affective decision making operates independently of the conscious, deliberate looping mechanisms. The latter are important and powerful; yet, they are only secondary to the core neural mechanism of affective decision making we have identified. Thus, as Zajonc (1980) put it, “preferences [often] need no inferences (page 151).”

7. CONCLUSIONS

7.1 Evaluating the Biosocial Model

In the present paper, we drew on recent advances in both neuroscience and animal behavior and proposed that when people face a conflict between two
competing behaviors such as approach versus avoidance and go versus no-go, they develop a new affective disposition. This newly emerging affect enables one to select a response while forming the basis for an elaborate cognition that justifies the selected response. This model, called the biosocial model, reconceptualized dissonance as affective (rather than cognitive). It explains both how decisions are made and how the decisions that are made are subsequently rationalized. Moreover, by postulating both top-down and bottom-up regulatory pathways, the model integrates all major moderating factors of dissonance, including action orientation, self-affirmation, mortality salience, and culture. Equally important, the model provides a coherent framework to understand a disparate set of affective and motivational phenomena, including animal behaviors that mimic effort justification, addiction, and ego-depletion. Furthermore, the model suggests why humans are affectively and motivationally attached to symbols of culture. The mechanism stipulated by the biosocial model may be one significant biological preadaptation for symbolic culture.

The model is grounded in known brain mechanisms. Although, at this point, not all details are supported by existing data, they are testable with currently available neuroimaging methods such as EEG and fMRI. By specifying the underlying brain mechanisms, the model makes it possible to locate the mechanisms of affective decision making within a broader, expanding theoretical framework of affective, cognitive, and social neuroscience (Lieberman, 2007). Moreover, the model is also socioculturally open. In particular, the regulatory pathways are likely affected by sociocultural learning and conditioning. Numerous sociocultural inputs, including action versus state orientation, religion, self-affirmation, and images of the Meadian generalized other, can systematically modulate the basic decision-making mechanics in a highly systematic fashion. Thus, the model can be legitimately called “biosocial.” It can illuminate, not only the underlying neural mechanisms, but also the nature of sociocultural reinforcement contingencies that shape these mechanisms.

7.2 How Cognitive is Cognitive Dissonance?

As noted earlier, dissonance has long been considered cognitive in the sense that inconsistencies among beliefs are considered to define dissonance. This perspective was instrumental in establishing cognitive theories in social psychology during the 1950s and 1960s, thereby paving the way to the productive era of social cognition research in the 1970s and 1980s (Fiske & Taylor,
The biosocial model, in contrast, suggests that the conflict at issue is not purely cognitive. It is much more behavioral, such as conflict between approach and avoidance or between go-response and no-go response. This consideration may shed some new light on the question of why not all cognitive beliefs lend themselves to dissonance. Some beliefs give rise to behavioral tendencies more readily or more vigorously than others, and according to the biosocial model, only those cognitions with direct implications for behavior lead to dissonance.

For example, in most contemporary secular societies, knowledge that the earth rotates around the sun is completely disconnected from behavioral reality. Thus, even though it clearly contradicts everything else people experience about the relationship between the earth and the sun, this conflict does not produce dissonance. Imagine, however, what might have happened to someone who practiced science at the dawn of Modern Europe like Galileo Galilei did in seventieth century Italy. With his commitment to the then-emerging science, the belief about the rotation of the earth would have carried entirely different behavioral implications. Galileo must have felt tempted, and perhaps obligated in the name of science, to express a view that was entirely incompatible with, and thus heretical to, a central dogma of the prevailing moral authority of the time, the Catholic Church. He must have been torn apart by dissonance (which, we should add, must have made his commitment to science even firmer than ever). This example illustrates why cognitions are important only to the extent that they have power to mobilize potent behavioral responses. Otherwise, cognitions are irrelevant vis-à-vis dissonance, and accordingly, the role they have in dissonance is inevitably secondary.

### 7.3 Future Directions

The biosocial model offers some important directions for future research. First, it is important to test specific neural mechanisms and pathways involved in certain key social psychological processes, such as action orientation, self-affirmation, and mortality salience. The biosocial model is specific enough to guide and direct such research in a systematic fashion. For example, self-affirmation should downregulate dACC through the midbrain DA pathway. Whether this is valid or not is consequential to cognition, emotion, and behavior in many social situations. Hence, empirically addressing this and other related questions is important for the future expansion and elaboration of social psychological theories.
Second, the biosocial model opens the door to an uncharted terrain of social genomics (Cole, 2009; Kim & Sasaki, 2014). For example, dissonance is fundamentally affective and motivational. Moreover, much of the striatal affective mechanism that is placed at the core of the dissonance process is innervated strongly by dopamine (DA) neurons (Berridge & Robinson, 2003). Accordingly, there is every reason to believe that some DA system genes are involved in modulating the functioning of core mechanisms of dissonance. One gene that is often consequential in DA-related processes such as reward processing and reinforcement learning is the dopamine D4 receptor gene, *DRD4* (Kitayama, King, et al., 2014). This gene, however, is linked to other genes that are relevant to the DA system including other DA receptor genes such as *DRD2* and *DRD3*, DA transporter genes (e.g., *DAT1*), as well as those involved in dopamine clearance (*COMT, MAOA*, and *MAOB*) (Saez, Set, & Hsu, 2014). By explicating the functions of these genes for neural processing and behavior, either alone or in combination, it may become possible to triangulate the nature of relevant neural mechanisms in greater detail.

In closing this paper, we wish to point out that the affective mechanisms articulated by the biosocial model are likely to play fundamental roles in maintaining human society and culture. To be a cultural member is to be conditioned, affectively and motivationally, to a system of signs established in a particular, geographically, religiously, and historically constituted group. In this sense, humans may all prove to be Robinson’s “sign trackers” to varying degrees. The underlying biology of reward processing may be crucial in generating a motivation for people to commit themselves to their traditions, religions, nations, and other signs of their culture. From this point of view, it should not come as a surprise that certain DA system genes are closely associated with the acquisition of cultural worldviews (Kitayama, King, et al., 2014). The biosocial model, then, may be pointing to new avenues of research in cultural psychology and beyond by bringing to the foreground hitherto neglected issues and questions about the biological basis of culture and society.

**ACKNOWLEDGMENTS**

We thank Nobuhiro Abe, Kent Berridge, Keisei Izuma, Hazel Markus, Richard Nisbett, Terry Robinson, and members of the University of Michigan Culture and Cognition Lab for their helpful comments on an earlier draft. Research reported here was supported by grants from National Science Foundation (BCS 0717982 and SES 1325881) and National Institute for Aging (5RO129509-02).
REFERENCES


Biosocial Model of Affective Decision Making

Author's personal copy


Immordino-Yang, M. H., Yang, X.-F., & Damasio, H. (2014). Correlations between social-emotional feelings and anterior insula activity are independent from visceral states but influenced by culture. *Frontiers in Human Neuroscience, 8*.


