Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Social norms, self-control, and the value of antisocial behavior

Joshua W Buckholtz[1,2,3]

Social norms facilitate large-scale cooperation by promoting prosocial interactions and constraining antisocial behavior. Dominant models of norm compliance emphasize the role of effortful, capacity-limited inhibitory control in prosocial cooperation. Similarly, clinical science has focused on inhibitory deficits as a key source of persistent norm-violating behavior. Support for an inhibition-based 'braking success/braking failure' (BSBF) model is derived from evidence of dorsolateral prefrontal cortex (DLPFC) engagement during norm-guided behavior, and of DLPFC dysfunction in antisocial individuals. However, three challenges motivate an alternative explanation for links between self-control, DLPFC, and norm-based behavior. Here, I propose a value-based alternative to the BSBF model, in which prosocial norm compliance and antisocial norm violations both arise from interactions between prefrontal model-based and striatal model-free decision-making systems.

Addresses
[1] Department of Psychology, United States
[2] Center for Brain Science, Harvard University, United States
[3] Department of Psychiatry, Massachusetts General Hospital, United States

Corresponding author: Buckholtz, Joshua W. (joshua_buckholtz@harvard.edu)

Current Opinion in Behavioral Sciences 2015, **3**:122–129

This review comes from a themed issue on **Social behavior**

Edited by **Molly J Crockett** and **Amy Cuddy**

For a complete overview see the Issue and the Editorial

Available online 3rd April 2015

http://dx.doi.org/10.1016/j.cobeha.2015.03.004

Our social landscape is shaped by social norms, a set of prescriptive and proscriptive rules that comprise the 'grammar of social interaction' [1] for *Homo sapiens*. Injunctions against dishonesty in exchange, physical harm, and theft promote social stability, community peace and economic prosperity. Thus, many have argued that norms were key for enabling the ultrasociality that is a signature of our species. In turn, their ability to facilitate cooperation depends on the widespread maintenance of norm compliance, accomplished by sanctioning norm violations [2] and by rewarding norm consistent (prosocial) behavior [3]. However, while norm compliance is extensive, it is far from universal, and our understanding of the mechanisms that drive norm violations (antisocial behavior) remains limited. Given the estimated societal costs of antisocial behavior — upwards of $1 trillion annually, by some accounts [4] — our relatively poor insight into its biological origins impedes treatment development and limits our ability to make informed policy decisions.

Social norms and self-control

Norm compliance often involves engaging in other-regarding behavior that is counter to an agent's immediate selfish interests. Thus, many have argued that norm-based behavior requires 'self-control,' [5,6] a term that is used somewhat interchangeably with 'self-regulation,' 'impulse control,' 'cognitive control,' and 'executive function'. If prosocial behavior requires effortful self-control, it would thus naturally follow that persistent norm violating (antisocial) behavior arises from self-control failure, manifest as an inability to appropriately inhibit self-interested decision-making. Indeed, this is the conceptual foundation for a highly influential theory of crime (the 'Self-Control Theory of Crime') [7], and clinical science research has largely focused on identifying the cognitive and neural bases of self-control deficits in antisocial individuals [8].

The 'brakes' metaphor of self-control

The prevailing conceptualization of self-control in studies of norm-based cooperative behavior is that of an effortful, capacity-limited resource, used to inhibit automatic or prepotent responses. Social psychologists and behavioral economists have argued that cooperation arises from the capacity to actively override selfish impulses in order to promote selection of an alternative norm-consistent choice option [5,6]. Similarly, clinical scientists have proposed that antisocial behavior results from a deficit in the capacity to actively inhibit the execution of prepotent responses to threat and/or reward associated stimuli [8,9]. Together, these two perspectives comprise a dominant viewpoint wherein 'braking success' produces prosocial behavior and 'braking failure' leads to antisocial behavior. While this 'braking success/braking failure' (BSBF) model of norm-guided behavior has both intuitive appeal and experimental support, recent work suggests that the BSBF model is incomplete. In what follows, I will review brain imaging findings that are often cited to support the BSBF model, highlight conceptual and empirical challenges to the model, and articulate an alternative framework grounded in decision science.

Brakes, localized?: dorsolateral prefrontal cortex

As noted above, many have suggested that norm-based prosocial behavior involves the deployment of self-control to actively inhibit automatic self-interested responses. Neurobiological studies of norm-based behavior have largely adopted this conceptualization, identifying the locus of the self-control 'braking' system in lateral prefrontal cortex. For example, dorsolateral prefrontal cortex (DLPFC) activation has been consistently noted in fMRI tasks that assess prosocial ('altruistic') punishment of norm violations, as well as in studies of cooperation, fairness, and social norm compliance [2,10]. A recent meta-analysis suggests that DLPFC activity during norm-based decision-making reflects the need for 'cognitive control from a reflective and deliberate System 2 to resolve conflict by ... over-riding self-interest' [11]. Of note, disruptive brain stimulation to DLPFC reduces both prosocial norm-enforcement [12] and norm compliance [13••]. As other studies have shown that this region is important for response inhibition, some have inferred that the ability to actively override selfish or otherwise maladaptive responses is a cognitive *sin qua non* for prosocial behavior [5].

Similarly, it is widely assumed that antisocial behavior results from impaired inhibitory control [8,9]. This perspective is supported by behavioral and neurobiological evidence that antisocial individuals show deficient cognitive control in the context of heightened reactivity to threat and/or reward cues. Antisocial offenders exhibit reduced gray matter volume and cortical thickness within DLPFC [8,14], as well as compromised DLPFC activation during classic neuropsychological indices of inhibitory control [14,15]. By contrast, antisocial individuals appear to have relatively exaggerated responses to threat stimuli (within the amygdala) and reward cues (within the striatum) [16••,17–19]. Interpreted through the lens of the BSBF model, such findings are taken as evidence that antisocial behavior occurs when bottom up 'affective' signals activate or generate a prepotent behavioral response that is inadequately inhibited by top down 'cognitive' resources due to poor prefrontal control.

Taken together, a large body of work provides convergent support for the idea that DLPFC is causally involved in norm compliance, and that antisocial individuals exhibit deficits in DLPFC structure and function. While such findings appear to fit the BSBF model, three key challenges suggest that an alternative explanation for the link between DLPFC and pro-/anti-social behavior merits consideration.

Challenge 1: prosociality, no brakes required

The first challenge comes from recent findings that prosocial behavior does not necessarily require the controlled inhibition of self-interested or maladaptive automatic responses. Cooperation in both laboratory and real-life contexts is highest under conditions that promote fast, intuitive responding [20]. Developmental and cross-species data offer further support for the intuitive, automatic nature of prosociality: chimpanzees — even very young ones — engage in costly helping behavior, and prosocial responding in human children emerges prior to the development of inhibitory control [21]. Finally, neuroimaging work suggests that costly prosocial behavior may be facilitated by the heightened value assigned to prosocial choice options relative to available alternatives, rather than the active inhibition of automatic antisocial responses [22]. Together, these findings imply that norm-consistent behavior does not necessarily require effortful inhibition.

Challenge 2: there is no such (single) thing as 'self-control'

The construct of self-control has a deceptively complex latent architecture. There is overwhelming evidence for multidimensionality, with several distinct cognitive components that can be grouped into at least two broad domains [23] (Figure 1). Capacities encompassed by the domain of 'response inhibition' enable agents to use internal representations or external cues to inhibit prepotent motor responses. A second domain contains processes that promote adaptive choice behavior by estimating the subjective value of different choice options and selecting actions based on optimal utility. Lesion and drug studies in animals, along with factor analytic work in humans, converge to suggest that response inhibition and value-based decision making are largely distinct cognitive capacities [24••] with dissociable neurobiological substrates [23]. Despite the evident heterogeneity of 'self-control,' it is often conflated with response inhibition or assessed by trait measures that map poorly to identifiable cognitive processes [25]. The net result of this construct neglect is two-fold. First, research on the relationship between self-control and norm-guided behavior has largely focused on response inhibition, while other facets of self-control remain relatively unexplored. Second, DLPFC engagement (or dysfunction) is often presumed to reflect inhibitory control, leaving alternative explanations for the role of DLPFC in pro-social and anti-social behavior relatively unexamined.

Challenge 3: DLPFC reconsidered

Early work on the role of DLPFC in adaptive behavior was largely focused on inhibitory control: lesion, electrophysiology, and functional imaging studies demonstrated that DLPFC is crucial for the ability to use goals, task rules and external cues to inhibit the execution of contextually inappropriate motor responses [26–32]. However, recent brain imaging and electrophysiological data have broadened the view of DLPFC to include a crucial role in value-based decision-making. Lateral PFC has been shown to be important for integrating value-related