Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

When is a loss a loss? Excitatory and inhibitory processes in loss-related decision-making

Ben Seymour[1,2,3,4], Masaki Maruyama[2,3] and Benedetto De Martino[4]

One of the puzzles in neuroeconomics is the inconsistent pattern of brain response seen in the striatum during evaluation of losses. In some studies striatal responses appear to represent loss as a negative reward (BOLD deactivation), while in others as positive punishment (BOLD activation). We argue that these discrepancies can be explained by the existence of two fundamentally different types of loss: *excitatory losses* signaling the presence of substantive punishment, and *inhibitory losses* signaling cessation or omission of reward. We then map different theories of motivational opponency to loss related decision-making, and highlight five distinct underlying computational processes. We suggest that this excitatory–inhibitory model of loss provides a neurobiological framework for understanding reference dependence in behavioral economics.

Addresses
[1] Computational and Biological Learning Laboratory, Department of Engineering, University of Cambridge CB2 1PZ, United Kingdom
[2] Center for Information and Neural Networks, National Institute for Information and Communications Technology, 1-4 Yamadaoka, Suita City, Osaka 565-0871, Japan
[3] Immunology Frontier Research Center, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japan
[4] Department of Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

Corresponding author: Seymour, Ben (bjs49@cam.ac.uk)

Current Opinion in Behavioral Sciences 2015, **5**:122–127

This review comes from a themed issue on **Neuroeconomics**

Edited by **John P O'Doherty** and **Colin C Camerer**

For a complete overview see the Issue and the Editorial

Available online 3rd October 2015

http://dx.doi.org/10.1016/j.cobeha.2015.09.003

Introduction

Over the past decade a set of divergent observations have emerged in human neuroimaging studies of monetary loss. In studies of the receipt (or prospect) of financial loss, neuroimaging responses sometimes exhibit deactivation in BOLD signal of striatal brain areas associated with motivation and decision-making (caudate, putamen, and nucleus accumbens) [1•,2,3,4•], or little change at all [5]. This has often been seen as consistent with a primary role of these regions in reward-related processing, and these negative responses are usually seen in the same regions showing activation to monetary gains. With emerging evidence that striatal BOLD responses to reward were well described by prediction error activity in the context of passive prediction tasks (Pavlovian learning), it was generally assumed that this activity represented a single reward-specific and putatively dopamine-related signal [6•,7].

However, this theory suffered when other studies involving loss, and especially involving primary punishments such as pain, revealed positive activation in the striatum, in very similar regions to those that showed *deactivations* to financial loss [8,9]. Furthermore, the pattern of activity resembled a prediction error, just like a reward prediction with the opposite sign. This suggested that either the striatum was encoding a more complex signal than originally thought — perhaps some sort of selective salience signal [10], or that there was a second system for encoding aversive outcomes that comes online with physical, but less so financial, punishment. Why financial loss might less reliably activate this system was unclear, but one could posit it might be related to the fact that physical punishments are primary outcomes 'consumed' immediately, whereas money is a secondary outcome whose real outcome is fulfilled at a later date. A more reliable way to 'activate' the striatum to loss was introduced with a clever design from Delgado and colleagues: they had subjects begin the experiment with a task in which subjects could earn a decent sized money pot. Then in a second, seemingly unrelated experiment, they underwent a loss-conditioning study, which revealed positive activation to monetary loss [11•]. This result, together with a more recent one [12], suggest that losing money that had been earned on a previous task in some way rendered it sufficient to reliably activate positive aversive coding.

This raises the question as to what makes a loss look sometimes primarily like a negative reward, and at other times like a positive punishment. This is important, because if there are substantially different ways of representing losses in the brain, then the associated loss behavior may have very different characteristics.

To make matters more complex, subsequent studies involving the capacity to make active choices over monetary loss or pain, that is, reducing or avoiding punishment, did not fit either pattern simply. Here, for both money *and* pain, striatal activity shows positive activation for avoidance actions and avoided outcomes [13,14]. Rather than representing the magnitude of the expected punishment (in probabilistic avoidance), it seemed to represent the relative positive value of avoidance [15]. That is, activity again looks like a reward signal — this time for actions, with no consistent evidence of an aversive striatal system in operation, even for painful outcomes. A positive aversive signal is sometimes seen elsewhere, such as the anterior insula cortex [16], but its contribution to decision making was less clear.

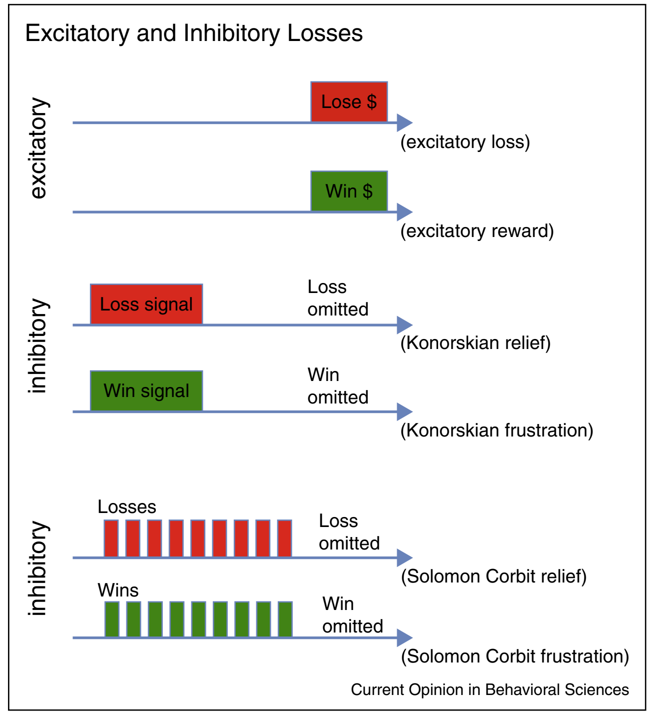
Excitatory and inhibitory loss

Animal learning theory provides a structured approach to understanding the relationship between gains and losses, and there is good evidence of the existence of two separate motivational pathways for outcome prediction: one governing rewards, and the other governing punishments [17]. In particular, accounts of interaction between the two systems yield two distinct types of punishment: excitatory, and inhibitory, depending on the context that defines the nature of punishment. Accordingly, inhibitory values emerge from two different instances for appetitive-aversive opponency: omission (Konorskian [18]) opponency, and offset (Solomon–Corbit [19]) opponency (Figure 1).

Omission opponency describes the frustrative loss that occurs when an expected reward does not occur. Here, excitatory losses are due to the positive presence of a punishment, and inhibitory losses due to absence of an expected gain. A slightly different type of frustrative loss occurs when a tonically presented reward terminates. In this case, Solomon and Corbitt proposed the accrual of a slow adaptive process from which acute changes were compared. Both processes illustrate the clear distinction between excitatory and inhibitory losses, with the inhibitory type being generated either comparison of neutral outcome with an expectation of or tonic baseline level of reward. This is exactly mirrored in the opposite valence: inhibitory reward being evoked with the relief at the termination or omission of punishment [20].

The existence of different types of loss offers an explanation into the pattern of brain responses seen above. In most experiments, for ethical and practical reasons, loss is operationalized by a reduction in the participant monetary compensation (future expected reward). This procedure would augment the inhibitory loss representation and therefore tend toward a deactivation in striatal areas. However, for primary punishments and financial outcomes that were already considered 'owned', we would expect a dominant excitatory loss representation, and activation of an aversive system observable in the striatum. In many situations, it may be that excitatory and inhibitory processes co-occur [21], and hence partially or fully cancel out the subsequent fMRI BOLD response.

Figure 1



Excitatory and inhibitory processes underlying reward and punishment. Excitatory values occur with the receipt, or prediction of receipt, of a primary reward or punishment. Inhibitory values occur with either the omission of an expected outcome (e.g. requiring, of course, an expectation to be generated by some process, such as Pavlovian Conditioning), or with the termination of a tonic or repetitively received outcome.

From loss prediction to decision-making

How then, is loss-related computation related to decision-making? Clearly, what motivates choice in the context of any type of loss is a desire to reduce it, and this can be used to *define* loss or punishment. The Konorskian and Solomon–Corbitt framework deals with passive (Pavlovian) predictions, but are conventionally thought to govern two distinct types of aversive decision: avoidance, and escape [22,23]. The control of escape and avoidance may be different, because the nature by which information about outcomes is garnered is different, but in both cases it is the absence of punishment that motivates behavior. The paradox created by the ability of 'nothing' to act as an incentive and reinforce actions has stimulated considerable research and debate [24]. Two putative solutions to the avoidance problem are provided by inhibitory *rewards*: in the one case (two-factor theory), behavior is driven by the escape from fear (i.e. offset relief) elicited by any signal that predicts punishment [25]. In a second case (safety signal hypothesis), behavior is driven by the