Brain and Cognition 88 (2014) 55–64

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

# Neural correlates of social perception on response bias

Yeon Soon Shin, Hye-young Kim, Sanghoon Han *

*Department of Psychology, Yonsei University, Seoul 120-749, Republic of Korea*

ARTICLE INFO

*Article history:*
Accepted 27 April 2014
Available online 22 May 2014

*Keywords:*
Response bias
Social perception
Feedback-based learning
Caudate nucleus
DLPFC
fMRI

ABSTRACT

Accurate person perception is crucial in social decision-making. One of the central elements in successful social perception is the ability to understand another's response bias; this is because the same behavior can represent different inner states depending on whether other people are yea-sayers or naysayers. In the present study, we have tried to investigate how the internal biases of others are perceived. Using a multi-trial learning paradigm, perceivers made predictions about a target's responses to various suggested activities and then received feedback for each prediction trial-by-trial. Our hypotheses were that (1) the internal decision criterion of the targets would be realized through repeated experiences, and (2) due to positive–negative asymmetry, yea-sayers would be recognized more gradually than naysayers through the probabilistic integration of repeated experiences. To find neural evidence that tracks probabilistic integration when forming person knowledge on response biases, we employed a model-based fMRI with a State-Space Model. We discovered that person knowledge about yea-sayers modulated several brain regions, including caudate nucleus, DLPFC, hippocampus, etc. Moreover, when person knowledge was updated with incorrect performance feedback, brain regions including the caudate nucleus, DLPFC, dmPFC, and TPJ were also involved. There were overlapping regions for both processes, caudate nucleus and DLPFC, suggesting that these regions take crucial roles in forming person knowledge with repeated feedback, while reflecting acquired information up to the current prediction.


## 1. Introduction

A substantial part of our lives consists of meeting other people and getting to know them better. To thrive in our social life, we spend considerable time speculating on what others would think and do (Dunbar, 2003). That is because having accurate knowledge about other individuals is a key to success in this area (for a review, see Zaki & Ochsner, 2011). In this sense, it is crucial to know the degree to which people's behavior genuinely represents their mind. Response biases are closely related to this representational discrepancy between observable behavior and internal states of mind. According to classical decision theories (Green, 1966; Yonelinas, 2002), the criterion is an important factor that determines behavior, because a decision results from an interaction between the criterion and evidence (i.e. the strength of stimuli). A response bias entails that the decision criterion is biased, and in this manner, evidence falls above the criterion with either a high or a low probability. Thus, those who have a liberal criterion (*yea-sayers*) are likely to give positive responses, while those with a more conservative criterion (*naysayers*) are less likely to do so. Interestingly, response biases are stable across different contexts (Berg, 1955; Couch & Keniston, 1960; Furnham, 1986), and reflect underlying personality traits (e.g., acquiescence, agreement, and social desirability) (Couch & Keniston, 1960; Furnham, 1986). Thus, response biases carry valuable information required for understanding others' current behavior and making correct predictions about their future behaviors as well. Despite its importance in successful social cognition, however, little is known about how we perceive others' response biases in social interactions and which neural regions are involved in this process.

The key question that needs to be answered first is how we come to realize others' response biases. A growing body of research have shed light on the neural underpinnings for person impression formation and the update process with inconsistent information which violates the first impression (Bhanji & Beer, 2013; Ma et al., 2012; Mende-Siedlecki, Baron, & Todorov, 2013; Mende-Siedlecki, Cai, & Todorov, 2013). Although these studies neatly showed how rapidly-formed impression is updated with inconsistent evidence, in the current study, we explored further to investigate how impression is gradually formed over the multiple encounter with the opposite person. Unlike in a spontaneous trait inference, a perceiver should integrate a number of incidents in order to read

* Corresponding author.
*E-mail addresses:* ys.cecilia.shin@gmail.com (Y.S. Shin), hyeyoung.astin.kim@gmail.com (H.-y. Kim), sanghoon.han@yonsei.ac.kr (S. Han).

http://dx.doi.org/10.1016/j.bandc.2014.04.012
0278-2626/

how frequently the person answered positively/negatively. Thus, the crux of this problem lies in the multiple experiences with the person and the way we attribute those experiences. A single response observed only once is not attributable to either a situational factor or an internal response bias, and this fact necessitates multiple observations. Moreover, it is still impossible to infer a response bias if attribution is made to an external context. Consequently, we should focus more attention on the internal state of the person. In this manner, the multiple responses of a person can be generalized and integrated, and thus, can contribute to refining our knowledge about a single response criterion. In the repeated experiences, spontaneous probabilistic computation is likely to take place, like in typical multi-trial feedback-based learning (for a review, see Niv, 2009). That is, response observation would not necessarily explicitly bring up the concept of response bias, but the experiences will instead be calculated probabilistically and stored for future use. To quantify the amount of such probabilistic knowledge (i.e., learning state), a learning model named State-Space Model (SSM, Smith et al., 2004) is used. Previous studies on SSM showed that this model is more sensitive than other Reinforcement Learning (RL) models in capturing hidden learning performance (i.e., the degree to which knowledge is formed). This is because the model assumes an ideal observer who knows the entire trials when fitting observed data into a hidden learning equation (Kakade & Dayan, 2002; Smith et al., 2004) while other RL model only considers the trials up to the current observation. Therefore, SSM has strong validity in that the estimated amount of accumulated knowledge that is obtained from the model is well tracked at a neural level (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Smith et al., 2004; Solomon, Smith, Frank, Ly, & Carter, 2011).

In addition to accumulated knowledge, knowledge-updating process itself is worth examining as well. Incorrect performance feedback is especially important here, because it elicits internal expectation violation, and so guides alternative correct predictions (Holroyd & Coles, 2002; Zanolie, Van Leijenhorst, Rombouts, & Crone, 2008). Although learners may capitalize on both correct and incorrect performance feedback, correct feedback conveys no more information than has already been accrued. In this sense, negative outcomes (i.e., "wrong") in feedback-based gradual learning have greater informational value than positive outcomes (i.e., "right").

Positivity and negativity of the biases is another critical issue. Given that yea-sayers have a higher probability of giving *positive* responses, while naysayers have a higher probability of giving *negative* ones, the positivity and negativity of responses would affect the way repeated experiences are generalized into knowledge about criterion. If positivity and negativity exhibit an asymmetrical influence upon the generalization process, this can have two possible consequences. The first is that positive responses are more readily generalizable and so serve as a better means of highlighting the underlying response criterion. In this way, observing a "yes" response would contribute more to person knowledge (asymmetrical integration – positivity dominance). The second possibility is that it is easier to recognize and integrate knowledge about a person's decision criterion from their negative responses (asymmetrical integration – negative dominance). On the other hand, if positivity and negativity do not asymmetrically influence the generalization process, observing a "yes" or a "no" would equally develop into adequate person knowledge about yea-sayers and naysayers (symmetrical integration).

In line with potential asymmetrical integration – in particular, the positivity dominance hypothesis – a substantial body of literature has shed light on positive–negative asymmetry in a range of diverse cognitive domains, such as valuation (Kahneman & Tversky, 1979), mood (Forgas, 1998), and episodic memory (Kensinger & Schacter, 2006; Ochsner, 2000). For example, it was discovered from mood-induced processing differences (Bless, Mackie, & Schwarz, 1992; Bless et al., 1996) that a perceiver is more likely to commit fundamental attribution errors when they are in a good mood, so that they tend to attribute the behavior of others to dispositional factors, while neglecting the role of situations (Forgas, 1998). More importantly, positivity itself also plays a role in the degree to which representation is generalized. As Tolstoy observes in a famous statement from *Anna Karenina*, "Happy families are all alike; every unhappy family is unhappy in its own way" (quoted in Unkelbach, Fiedler, Bayer, Stegmuller, & Danner, 2008), there seems to be much less variety in positivity than there is in negativity. Supporting this observation, positive objects have denser semantic nodes and a more homogeneous representation, while negative objects have a more heterogeneous representation (Unkelbach et al., 2008). Similarly, there are lines of research that suggest positivity induces broader and more generalized cognitive processing. Positive mood expands attentional breadth (Fredrickson & Branigan, 2005) and increases exploration behavior (Fredrickson, 2001). Moreover, positivity also plays a role in memory. Memories about positive stimuli are less accurate (Ochsner, 2000), while, in contrast, those about negative stimuli are more detailed and accurate (Kensinger & Schacter, 2006). For example, Ochsner (2000) found that individuals respond that they "know", but do not "remember", the positive item, suggesting that people have a less detailed memory about the positive items they encounter. On the other hand, Kensinger and Schacter (2006) have demonstrated that our memories about negative episodes are formed in a more detailed manner.

In the current study, we sought to investigate how multiple responses are generalized into the concept of a response criterion and how we learn about positive and negative response biases to different degrees. To examine this process, we employed a feedback-based learning paradigm (Gluck, Shohamy, & Myers, 2002; Knowlton, Mangels, & Squire, 1996; Maddox, Ashby, Ing, & Pickering, 2004). In our experimental paradigm, participants made a prediction and observed other people's responses to various suggested activities. A respondent's answers were expected to serve as a cognitive feedback for the observer, who will then accumulate this information and generalize it in order to make predictions. Although similar to traditional weather prediction tasks (Knowlton et al., 1996), our paradigm is distinct in that the objects (i.e., the activity that a responder was asked to perform) varied trial-by-trial, with a target person and question ("Would she perform the activity?") fixed. By doing so, we focused on inducing generalized knowledge rather than activity-reaction associations. With functional magnetic resonance imaging, we further aimed to explore the neural correlates of both the representation of probabilistic knowledge in the brain and the knowledge update process. Furthermore, by using a conjunction analysis, we sought to locate the knowledge-updating regions of the brain that are modulated by previously acquired knowledge. An information-sensitive caudate nucleus and DLPFC were hypothesized as providing the means by which knowledge about response biases was updated while being modulated by the amount of information.

## 2. Experiment 1: behavioral study

We first conducted a behavioral experiment in which participants made a prediction about a responder's reaction (i.e. "yes" or "no") and received feedback from multiple cases. With this feedback-based learning paradigm, we aimed to investigate if learning occurs in line with the responder's actual response tendency, and if the learning performances for yea-saying and naysaying are potentially asymmetrical.