Behavioral contagion during learning about another agent’s risk-preferences acts on the neural representation of decision-risk

Shinsuke Suzuki1,3, Emily L. S. Jensen2, Peter Bossaerts3,4, and John P. O’Doherty1,4

1Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; 2Faculty of Business and Economics, The University of Melbourne, Carlton, VIC 3010, Australia; 3Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC 3052, Australia; and 4Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125

Edited by Ranulfo Romo, Universidad Nacional Autónoma de México, Mexico City, Distrito Federal, Mexico, and approved February 23, 2016 (received for review January 7, 2016)

Our attitude toward risk plays a crucial role in influencing our everyday decision-making. Despite its importance, little is known about how human risk-preference can be modulated by observing risky behavior in other agents at either the behavioral or the neural level. Using fMRI combined with computational modeling of behavioral data, we show that human risk-preference can be systematically altered by the act of observing and learning from others’ risk-related decisions. The contagion is driven specifically by brain regions involved in the assessment of risk: the behavioral shift is implemented via a neural representation of risk in the caudate nucleus, whereas the representations of other decision-related variables such as expected value are not affected. Furthermore, we uncover neural computations underlying learning about others’ risk-preferences and describe how these signals interact with the neural representation of risk in the caudate. Updating of the belief about others’ preferences is associated with neural activity in the dorsolateral prefrontal cortex, underlies the modulation of risk-preference. Taken together, these findings provide a mechanistic account for how observation of others’ risky behavior can modulate an individual’s own risk-preference.

An individual’s attitude toward risk can exert a profound influence on his/her life in a wide array of contexts (1, 2). For example, risk-attitude governs an individual’s decision to purchase a safe asset or to invest in a risky stock (3). Moreover, a risk-seeking attitude can lead to an increased tendency toward behaviors leading to adverse outcomes such as drug-taking, unsafe sexual behavior, pathological gambling, and other potentially life-threatening pursuits; on the other hand, a risk-averse tendency can result in a reduced prospect of attaining the potentially high gains associated with the pursuit of risky options (4, 5). Given the importance of risk-attitudes in influencing everyday behavior, considerable research has been conducted on the factors influencing risky decision-making. For instance, there is substantial evidence that a number of extraneous variables such as the framing of a decision context in terms of losses or gains (6, 7), exposure to stressful life events (8, 9), and experiences of losses and gains (10) can modulate risk-preferences. Less studied however is the role of a contagion effect (11) in modulating risk-seeking/averse behavior. That is, it remains elusive how one’s risk-related behavior is influenced by observing the behavior of others. The role of contagion may be especially important for understanding how and why risky behavior can become manifest in a number of critical situations. For example, observing a peer’s risk-seeking behaviors might exert a profound influence on conspecifics (12), resulting in an increased tendency toward risk-seeking behavior, especially during adolescence. Furthermore, the tendency of financial markets to collectively veer from bull to bear markets and back again (13) could arise in part because of the contagion of observing the risk-seeking or risk-averse investment behaviors of other market participants.

There is considerable evidence that contagion or conformity can affect an individual’s belief and behavior (11, 14, 15). For instance, an individual’s decision-making, including risk-related choice, can be changed by observing other peoples’ behavior (14–18), and the behavioral shift is reflected in value-related neural activity (16, 17, 19). However, little is known about the computational mechanisms underlying the contagion effect.

Here, we aimed to provide a mechanistic account of how contagion from observing or learning from other agents modulates an individual’s own risk-related behavior by testing the following three hypotheses. First, the act of observing and learning from other agents will alter an individual’s own risk-preferences. Second, the behavioral change will be associated specifically with modulation of the neural processing of risk. That is, the neural representation of risk per se, will be modulated, resulting in an increased or reduced perception of risk for gambles. Third, an interaction between neural systems implicated in representing risk and learning about others’ risk-preference will capture individual differences in susceptibility to the contagion effect.

Results

Experimental Task. To test these hypotheses, we scanned 24 human participants using functional magnetic resonance imaging (fMRI),

Significance

Why do we on some occasions engage in risky behavior but not on other occasions? Here, we explore the neural mechanisms of one possible account: a contagion effect. Using neuroimaging combined with computational modeling, we show that if we observe others behaving in a risk-seeking/risk-averse fashion, we become in turn more/less prone to risky behavior, and the behavioral shift is specifically implemented via neural processing of risk in a brain region, caudate nucleus. We further show that functional connectivity between the caudate and the dorsolateral prefrontal cortex, a region implicated in learning about others’ risk-attitude, is associated with susceptibility to the contagion effect, providing an account for how our own behavior can be influenced through observing other agents.

Author contributions: S.S., E.L.S.J., P.B., and J.P.O. designed research; S.S. and E.L.S.J. performed research; S.S., E.L.S.J., and J.P.O. analyzed data; and S.S., P.B., and J.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. Email: shinsuke.szk@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600092113/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1600092113
Experimental task and basic behaviors. (A) Timeline of each trial. On Self trials, participants decide whether to accept or reject a gamble within 4 s (response time range, 0.29–3.88 s; mean ± SD, 1.62 ± 0.51; data from all of the participants are collapsed). If participants accept, they can gamble for a specific amount of money; otherwise, they can take a guaranteed $10. The reward probability and magnitude of the gambles are varied on every trial and are represented by a pie chart (size of the green area indicates the probability; and the digits denote the magnitude). On Observe trials, participants are asked to observe a choice of another person, the “observee.” On Predict trials, participants predict the choice that the observee would make (response time range, 0.78–3.90 s; mean ± SD, 1.79 ± 0.56). Note that instruction phases were presented only during the first trial in each block. (B) Overall schedule. Sessions 1, 3, and 5 include only Self trials, whereas sessions 2 and 4 contain all of the three trial types in a block-wise manner. Orange, blocks of Self trials; blue, blocks of Observe trials; green, blocks of Predict trials. White digits denote the number of trials in each block. (C) Participants’ risk-preference in Session 1 (orange) and the two observers’ preferences (blue, mean ± SD). Positive and negative values indicate risk-seeking and -averse, respectively.

while they repeatedly chose between a risky gamble and a guaranteed $10 (Self trials) (Fig. 1A), observed decisions of confederates (Observe trials), and predicted the confederates’ decisions (Predict trials). The set of gambles presented (Fig. S1A and B) was designed to decorrelate the risk (variance of reward) from the expected value of reward across trials (1).

The experiment consisted of five sessions, in which the three types of trials were presented in a block-wise manner (Fig. 1B): Sessions 2 and 4 involved all three types, whereas Sessions 1, 3, and 5 included only Self trials. The confederate (“observee”) for Observe/Predict trials was different between Session 2 and 4; one for Session 2 was risk-averse and the other for Session 4 was risk-seeking, or vice versa (Fig. S1B). We instructed participants that the choices they observed were made by a real person recorded from a previous experiment. In actuality, however, the observees’ choices were generated by computer algorithms, as in previous studies (14, 15, 17, 19, 20).

Basic Behavioral Results. Consistent with previous findings (21), before observing the others’ decisions, the majority of participants exhibited risk-averse behavior, although there were considerable individual differences (Fig. 1C). Here, we defined each participant’s risk-preference as the proportion of gambles accepted relative to the proportion accepted by the risk-neutral agent (i.e., positive/negative values indicate risk-seeking/averse, respectively), unless specifically mentioned otherwise. This simple model-free measurement was highly consistent with other prevalent model-based measurements based on utility functions (Fig. S1 E and F).

Behavioral Evidence for Contagion of Risk-Preference. Participants’ behavior showed an effect of contagion. That is, the participants’ risk-preference was shifted toward the observees’ (see Fig. 2A for an example participant who exhibited a clear effect). The degree of contagion, defined as positive when the participant conformed to the observee (see Fig. 2A and the legend for details), was significantly positive ($P < 0.01$; Fig. 2B). This effect was also confirmed to be significant for both the risk-seeking and -averse observees ($P < 0.01$ for both; Fig. 2C) and replicated in an independent behavioral experiment (Fig. S2). A closer examination revealed a trend that the effect was more prominent when participants’ own risk-preference was incongruent with the observee’s, but the distance between participants’ and the observees’ preferences did not parametrically covary with the degree of contagion (Fig. S3). Furthermore, the degree of contagion was not significantly correlated with the proportion of correct predictions in Predictions trials ($P > 0.05$; Fig. 2D), implying that the contagion was not primarily triggered by predicting the observees’ choices.

To exclude other accounts for the behavioral shift, we conducted additional analyses. First, we examined whether participants became more “rational” by observing the observees’ choices and found no significant evidence of a change in their rationality across sessions (Fig. S4 A and B and SI Text 1). Second, we confirmed that the behavioral shift cannot be explained by “regression to the baseline” (Fig. S4C and SI Text 1).

Next, we aimed to show that the shift in participants’ choices across sessions was better captured by a change in their risk-preferences rather than a change in their subjective judgments about the probabilities (7, 22, 23). To this end, we constructed two computational models, one with varying risk-preference across sessions and the other with varying probability-weighting, and compared their goodness-of-fit (SI Text 2). The analysis revealed that the first model provided the better fit, indicating that the behavioral shift was better explained by a change in risk-preference. Similar analyses also revealed that it is unlikely that participants simply biased utility (16) or choice probability (24) of gambling options, possibly by copying the observee’s tendency to take a gamble, without changing their own risk-preference (SI Text 3 and SI Text 4). These behavioral results together suggest that decision-making under risk can be altered by observing others’ decisions through the change of one’s own risk-preference.

Finally, to examine the extent to which the contagion effect depends on observing the behavior of another human vs. a non-human computer agent, we conducted an additional behavioral experiment in which participants observed/predicted choices of another human and a computer agent on different sessions (see SI Methods for details). The result indicates that the contagion effect is present for both the human and computer observers ($P < 0.01$ for both of the sessions, with no significant difference between the sessions; $P = 0.20$). These results suggest that the extent to which the observed agent is human or artificial does not affect the extent to which the contagion effect is manifest.

Neural Encoding of Risk. At the neural level, risk was represented in a dorsal part of striatum: the caudate nucleus (Fig. 2E and
Posterior predictive distribution (PPD) suggests that the participants’ preferences were consistent with the observed data. To confirm this, the study used a computational model to simulate the participants’ preferences and compare them to the observed data. The results showed that the participants’ preferences were well-matched to the model predictions, indicating that contagion played a role in shaping their preferences.

The findings of this study can be summarized as follows:

1. **Contagion of Risk-Preference**: The study found evidence of contagion in risk-preference among participants. This effect was observed across different sessions, indicating that participants are influenced by the risk-preferences of their peers.

2. **Neural Representation of Risk**: Activity in the caudate nucleus was significantly correlated with the trial-by-trial risk of the gambling option, suggesting that the caudate plays a role in representing risk during decision-making.

3. **Behavioral Learning**: Participants learned about the preferences of others through observation and imitation, which influenced their own risk-preferences. This learning effect was observed in both risk-averse and risk-seeking participants.

4. **Predictive Performance**: The study found that participants who learned about the risk-preferences of others were able to predict the outcomes of these behaviors more accurately than those who did not.

The study’s findings have implications for understanding how social influence affects decision-making and can be applied to various domains, such as finance, economics, and psychology. The use of computational models in this study highlights the potential of leveraging mathematical tools to study complex social phenomena.
Neural Correlates of Learning About Others’ Risk-Preference. To examine the neural substrates of learning about others’ risk-preferences, we searched for brain regions associated with updating the belief about others’ risk-preference, which is captured by the Kullback–Leibler divergence ($D_{KL}$) between the posterior and the prior (26). The belief-updating signal, $D_{KL}$, was found to correlate with activity in dorsolateral prefrontal cortex (dlPFC) as well as other regions including dorsomedial prefrontal cortex (dmPFC) and inferior parietal lobule (IPL) at the time of confirmation in Observe trials (Fig. 3C and Table S3; $P < 0.05$ whole-brain FWE corrected at cluster level). Notably, the right dlPFC activity was also significant ($P < 0.05$) under the whole-brain FWE correction at voxel level. Furthermore, we confirmed these activations survived even when potential confounds, such as the observer’s response time, decision, motor responses, were included into the regression analysis [GLM V-2 (SI Methods)]. These results suggest that updating of the belief in learning about others’ risk-preference may occur in the dlPFC, dmPFC, and IPL—regions previously implicated in learning about others’ reward structure and mental states (19, 27–30).

Functional Connectivity Between the dlPFC and the Caudate. Given the above results together with previous findings of the anatomical and functional connectivity between dlPFC and caudate (31, 32), we further reasoned that functional connectivity between the two regions would account for individual differences in the degree of behavioral contagion. To test this, we conducted a psychophysiological interaction (PPI) analysis on the caudate with a psychological factor, the timing of self-decision, and a physiological factor, the dlPFC activity (see SI Methods for details). This analysis revealed that functional coupling between the two regions at the time of self-decision (i.e., effect of the PPI term) was significantly correlated with the degree of behavioral contagion across participants (Fig. 3D; $P < 0.05$; see SI Text 6 for robustness check). More specifically, the PPI effect was negative and around 0 for those participants with lower and higher behavioral contagion effect, respectively (Fig. 3D), and the main effect of the dlPFC activity was significantly positive ($P < 0.01$). A possible interpretation of these findings is that, for those individuals who are less susceptible to the contagion effect, the default connectivity between dlPFC and caudate is suppressed at the time of decision-making for the self.

Discussion

The present study uncovers the computational process by which contagion effects arising from observing the behavior of another agent can influence one’s own decision-making under risk.

Contagion Modulates Human Risk-Preference. Behaviorally, we demonstrate that human risk-preference can be altered by a contagion effect and rule out alternative possibilities such as changes in subjective judgment about the probabilities (22, 23) or simple bias for/against gambling options (16, 24). In economics and finance, the idea that risk-preferences can be changed is still controversial, and it is difficult to exclude the possibility that the observed shift in behaviors merely reflects the change of something else such as beliefs in expected value, reward probability, or integrated utility (33). Our behavioral analyses provide evidence for the view that risk-preferences can indeed be altered.

Why did the participants change their risk-preference toward the others? In our experiment, the participants were monetarily incentivized to learn others’ risk-preference and predict their future choices. One possibility is that the learning with an explicit incentive for correct prediction leaked over to influence the participants’ own choices. However, we found no association between the degree of contagion and the prediction performance in our data. Furthermore, previous studies have demonstrated that in various contexts, contagion can occur by merely observing/learning others’ without any explicit incentive for prediction (14–17, 19). When taken together, the evidence suggests that the contagion effect is not contingent on the provision of an explicit incentive for prediction. An important direction for future studies will be to identify the specific contextual elements that gives rise to the emergence of a contagion effect on risk-preference (see SI Text 7 for further discussion).

We also found that contagion occurs with both human and artificial agents. In a behavioral experiment, we found that participants changed their risk-preference after observing not only human but also computer observers. This finding has an implication for finance. In modern financial markets, the use of algorithmic trading (trades generated by artificial intelligence) has become increasingly popular (34). Our finding implies that contagion of risk-preference can work in such markets and potentially play a critical role in financial bubble formation and collapse.

Risk-Preference Is Altered Through the Modulation of the Neural Processing of Decision-Risk. The neural representation of decision-risk in the caudate nucleus was found to be specifically modulated, in line with the behavioral shift in risk-preferences. On the other hand, representations of other decision-related variables such as expected value, reward probability, reward magnitude, and integrated utility signals, were unaffected by the behavioral shift. These results indicate that the neural representation of decision-risk per se is directly modulated by the contagion effect, consistent with the view that risk-preferences are altered through changes in risk perception.
Our finding that risk signals that can be used as an input to the decision process are encoded in the caudate nucleus stands in contrast to some other studies that have reported risk representations in ventral striatum as well as insular cortex (1, 35, 36). A key difference between those previous studies and the present study is that those previous studies probed the neural representation of “anticipatory risk,” in that they measured activity during an anticipatory phase in which an outcome was imminent, without a choice being rendered or after a choice was made. In the present study, we designed our experimental task to capture neural processing of risk at the time of decision-making without the contribution of other risk-related effects such as anticipation risk and the effects of learning from reward feedback. The outcome of each choice was not revealed (37), ensuring that an outcome was not immediately anticipated and that valuations were not influenced by the history of previous outcomes (see (38, 39) for the discrimination of instructed and learned value information). According to our literature survey (Table S4), only a few studies used such a design, and they reported neural activity related to decision-risk in diverse regions including caudate and insula (37, 40–43). The present study provides additional evidence that caudate tracks decision-risk.

The contribution of the caudate to decision-risk is broadly consistent with the view that dorsal striatum is implicated in motivational and reward-related processes that involve decisions about action (44, 45). Combining this view with our finding that risk-related caudate activity remains even after controlling for motor-related responses, we suggest that the caudate is involved in risk-processing during decision-making, above and beyond any contribution of this structure in motor responding per se. Moreover, a number of clinical studies have demonstrated that individuals with anxiety disorders show reduced reward-related neural responses in caudate (46), while being more risk-averse than other clinical patients and normal control groups (47, 48). These findings broadly support our claims about a role for the caudate in risk representation and modulation of risk-preference. However, it is important to note that the present findings do not exclude the possibility that insular cortex and/or other structures could also play a role in encoding risk signals at the time of decision-making, above and beyond a role in anticipation risk, particularly under conditions where an outcome is imminent.

Biological and social sciences have accumulated evidence that human valuations can be altered by a contagion effect (14, 15, 17, 19, 49, 50). However, we know little about which components of the value computation is affected. Does human valuation work at the level of specific decision-related attributes or at the level of an integrated utility signal that combines across multiple types of attribute? Our findings in the context of decision-making under risk imply, in a broad sense, that human valuation can be modulated through a change in the neural processing of a particular decision-related attribute (in this case, the risk representation), rather than necessarily acting on an integrated utility signal.

Modulation of Risk-Preference Is Mediated by an Interaction Between Neural Systems Implicated in Representing Risk and Learning About Others. By using model-based analyses on the behavioral and fMRI data, we showed that the process of learning about others’ risk-preferences is well-captured by a Bayesian learning algorithm with the use of one’s own preference as a prior belief and that the belief-updating in the learning is associated with dIPFC. The behavioral finding suggests that an individual uses his/her own preference as a starting point for learning and making inferences about other people, compatible with the concepts of “anchoring-and-adjustment” (51) and “self-projection” (52, 53).

We also demonstrated via a connectivity analysis that functional coupling between the dIPFC and the caudate nucleus reflects individual differences in susceptibility to contagion. These results suggest that the contagion effect modulates the neural representation of risk in caudate through functional connectivity with the dIPFC. The role of a dorsolateral prefrontal–striatal circuit (31) in contagion of risk-preferences can be interpreted within a broader literate implicating this circuit in cognitive functions more generally such as in response-selection, planning, and set-shifting (32). The dIPFC and caudate have also been specifically implicated in goal-directed learning and model-based reinforcement-learning (54–56). Taken together, the present findings suggest that the contagion effect may act on neural circuits underpinning a model-based “cognitive” decision-making system.

The conclusions of the present study differ from previous findings by Chung et al. (16), who argued that contagion works by means of a constant bias to an integrated utility but not via risk-preference per se. Whereas the present study involved repeated opportunities to observe the risky behavior of another specific agent, in the Chung et al. study, participants interacted with multiple (six in total) anonymous agents throughout and thus did not have as much opportunity to learn about the specific risk-preferences of individual agents. Such task differences could potentially account for the difference in the overall findings between the studies. Our present findings indicate that when an individual has the opportunity to consistently observe the risky behavior of another agent, one’s own risk-preference can be directly influenced.

In conclusion, our results provide a computational account of how human risk-preferences are altered by the contagion effect. This finding has implications for economic and clinical studies. Although previous studies in economics and finance have linked financial bubbles to herd behaviors based on social learning (57, 58), our findings provide evidence for the existence of a novel and parsimonious effect that could potentially contribute to financial bubble formation and collapse: contagion of risk-preference through changes in the perception of risk. Furthermore, given that adolescent behavior can be strongly influenced by peers (12), contagion of risk-preference could play a significant role in leading to increases in adolescent risk-taking resulting in maladaptive behavioral outcomes.

Methods

This study was approved by the Institutional Review Board of the California Institute of Technology, and all participants gave their informed written consent. We provide a comprehensive description of the experimental procedures in SI Methods.

In Self trials, participants chose whether to “accept” or “reject” a gamble for themselves. If they chose accept, they gambled for some amount of money; otherwise, they took a guaranteed $10. Reward probability and magnitude of the gamble were varied in every trial (Fig. SI A), so that risk of the gamble (mathematical variance of reward) was decorrelated with the expected value of reward. In Observe trials, participants observed a choice made by the observer. The trials were designed to minimize differences from the Self trials, and so the timeline of a trial and the set of gambles presented were the same between the two types of trials (Fig. 1A and Fig. S1 A and B). Furthermore, in the participant instructions, we emphasized that the observers did not have any further information about the task such as knowledge of the outcome of the gambles. Predict trials were introduced to confirm that the participants learned the observer’s behavioral tendency (i.e., risk-preference) through the observation of his choices in Observe trials, and therefore the number of trials was less than the other two trial types (Fig. 1B and Fig. S1 A–C).

ACKNOWLEDGMENTS. We thank Simon Dunne and Keise Izuma for helpful comments on the manuscript. This work was supported by the Japan Society for the Promotion of Science (JSPS) Postdoctoral Fellowship for Research Abroad (S.S.) and the National Institute of Mental Health (NIMH) Caltech Conte Center for the Neurobiology of Social Decision Making (J.P.O.).


