

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

Shinsuke Suzuki<sup>a,1</sup>, Emily L. S. Jensen<sup>a</sup>, Peter Bossaerts<sup>b,c</sup>, and John P. O'Doherty<sup>a,d</sup>

<sup>a</sup>Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; <sup>b</sup>Faculty of Business and Economics, The University of Melbourne, Carlton, VIC 3010, Australia; <sup>c</sup>Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC 3052, Australia; and <sup>d</sup>Computation 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 (dlPFC). Functional coupling between the dlPFC and the caudate correlates with the degree of susceptibility to the contagion effect, suggesting that a frontal-subcortical loop, the so-called dorsolateral prefrontal-striatal circuit, 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.

risk | caudate | decision-making | conformity | neuroeconomics

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.

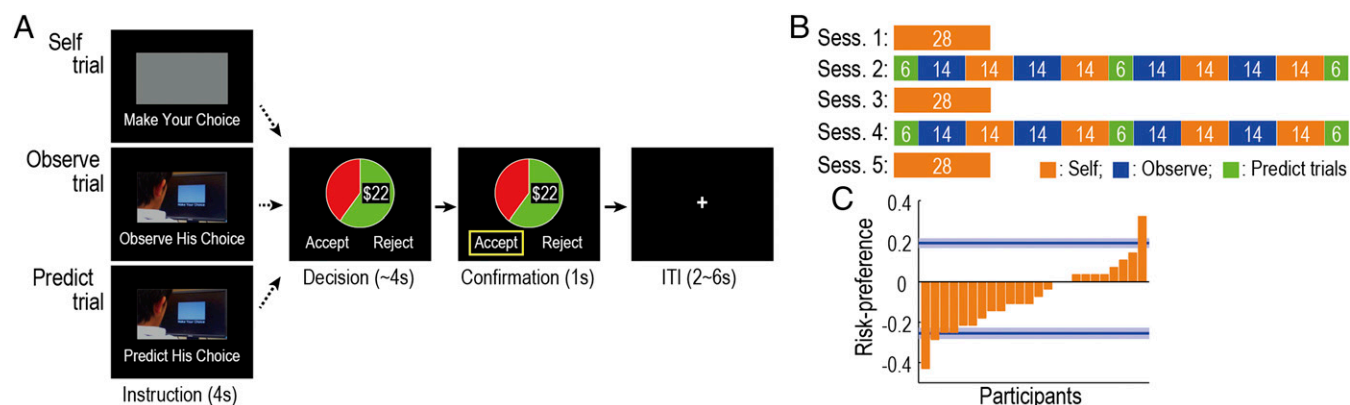
Author contributions: S.S., E.L.S.J., P.B., and J.P.O. designed research; S.S. and E.L.S.J. performed research; S.S., E.L.S.J., and J.P.O. analyzed data; and S.S., P.B., and J.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To 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](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1600092113/-DCSupplemental).



**Fig. 1.** 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  $\pm$  SD,  $1.62 \pm 0.51$ ; 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  $\pm$  SD,  $1.79 \pm 0.56$ ). 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 observees’ preferences (blue, mean  $\pm$  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. S1E 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

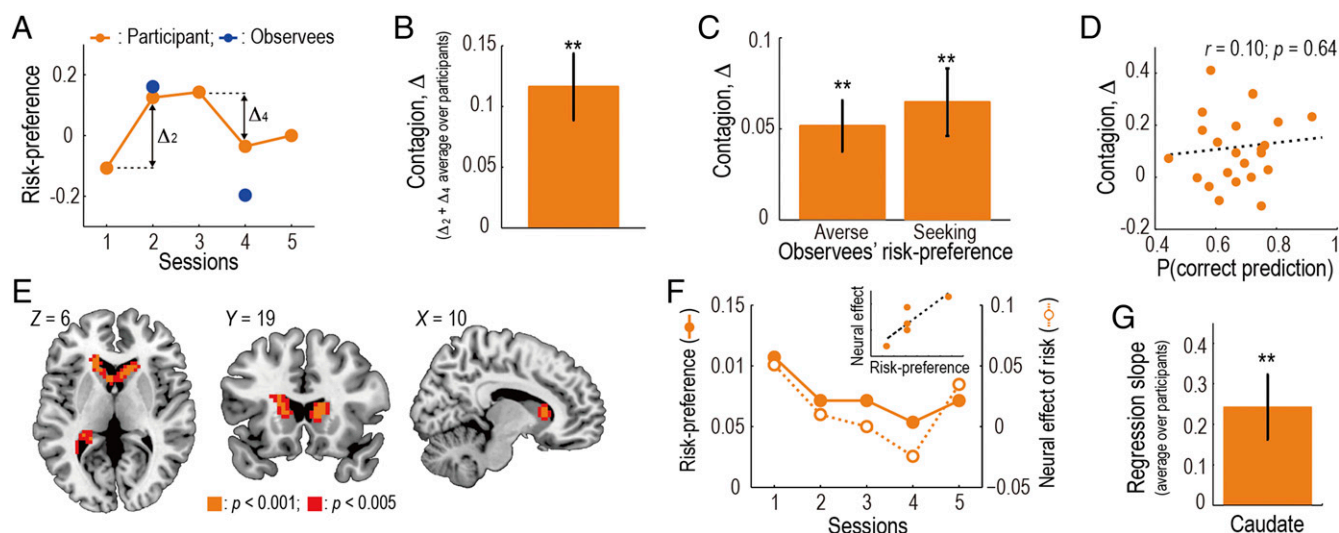
parametrically covary with the degree of contagion (Fig. S3). Furthermore, the degree of contagion was not significantly correlated with the proportion of correct predictions in Prediction trials ( $P > 0.05$ ; Fig. 2D), implying that the contagion was not primarily triggered by predicting the observees’ 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. S4A 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 observees ( $P < 0.01$  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



**Fig. 2.** Contagion of risk-preference. (A) Change of an example participant's risk-preference toward the observees'. Orange indicates the participant's risk-preference in each session, and blue indicates the observees' preference. The degree of contagion is defined as  $\Delta$ : the change of the risk-preference (i.e., proportion of gambles accepted) from the last session.  $\Delta$  is positive when the participant conformed to the observee; negative when she/he anticonformed. (B) Degree of contagion (mean  $\pm$  SEM across participants,  $n = 24$ ). The degree,  $\Delta$ , is defined as the sum of  $\Delta_2$  and  $\Delta_4$ .  $**P < 0.01$ . (C) Degree of contagion plotted separately for the risk-averse and -seeking observees (mean  $\pm$  SEM across participants). (D) Degree of contagion in each participant plotted as a function of the proportion of the correct predictions made on Predict trials. (E) Neural representation of risk. Activity in the caudate significantly correlated with risk of the gamble at the time of decision in Self trials [ $P < 0.05$  FWE corrected at cluster-level; GLM I (SI Methods)]. (F) Relation between the behavioral risk-preference and the neural effect of risk in the caudate in an example participant. Filled points indicate the participant's behavioral risk-preference in each session; open points indicate the neural effect of risk (i.e.,  $\beta$  value for the risk regressor in GLM I) in each session. (Inset) Scatter plot of the same data. (G) Relation between the behavioral risk-preference and the neural effect of risk (mean  $\pm$  SEM across participants). We regressed the neural effect in the caudate against the behavioral risk-preference across sessions and plot the average regression coefficient over the participants.

Table S1). Blood-oxygen-level dependent (BOLD) signal in the caudate significantly correlated with the trial-by-trial risk of the gambling option at the time of decision in Self trials [ $P < 0.05$  whole-brain familywise error (FWE) corrected at cluster level; general linear model (GLM) I (SI Methods)]. The right caudate activity was also significant ( $P < 0.05$ ) under the whole-brain FWE correction at voxel level. As a robustness check, we included the following variables into our regression model as regressors of no-interest (GLM I-2; SI Methods): the response time, the decision-related response, and the motor-related response. Running this revised model with the motor covariates on the anatomically defined caudate region of interest (25), we confirmed a significant effect of risk on the neural activity after adjusting for the effects of motor activity ( $P < 0.05$ ). The risk effect also continued to survive a whole-brain voxel-wise analysis with these motor covariates included, albeit at a reduced threshold ( $P < 0.005$  uncorrected; cluster size,  $k > 45$ ). Taken together, these results indicate that the caudate represents risk during decision-making, over and above effects of motor-responses and response-time.

**Neural Encoding of Risk Associated with the Behavioral Shift in Risk-Preference.** We then found that the neural effect of risk in the caudate covaried with the behavioral shift in risk-preference across sessions (see Fig. 2F for an example participant showing a clear effect). Here, the neural effect of risk in each session is expressed as a  $\beta$  value (i.e., regression coefficient) of the risk regressor from GLM I. To quantify this finding, we regressed the neural effect of risk in the caudate against the behavioral risk-preference across sessions and confirmed that the regression slope was significantly positive ( $P < 0.01$ ; Fig. 2G). The result does not change if we use the data from sessions including only Self trials (i.e., Sessions 1, 3, and 5) ( $P < 0.01$ ). We also conducted a whole-brain analysis to search for brain regions in which the neural effect of risk covaried with the behavioral preference

across sessions and found only one region: caudate nucleus ( $P < 0.05$  whole-brain FWE corrected at cluster level).

On the other hand, the expected value of the gambling option was correlated with activity in frontoparietal network brain regions including medial prefrontal cortex (mPFC) ( $P < 0.05$  whole-brain FWE corrected at cluster level; Table S1 and Fig. S5A). However, the neural effects of expected value were not correlated with the behavioral shift in risk-preference across sessions ( $P > 0.05$  for all of the activated clusters; see Fig. S5B and the legend for details). Moreover, we found that, consistent with a previous study (16), mPFC and other regions tracked the utility signal incorporating the change in risk-preference across trials, but the neural effects were not associated with the behavioral shift across sessions ( $P > 0.05$  for all of the activated clusters; Fig. S5C). We also confirmed that the behavioral change of risk-preference did not covary with the neural effects of other decision-related variables: reward probability, magnitude, or squared magnitude of the gambling option ( $P > 0.05$  for all of the activated clusters; Fig. S5D–F). These neural findings together support our hypothesis that contagion modulates the neural representation of risk per se, not the representations of other factors such as expected value or integrated utility.

**Learning About Others' Risk-Preference.** To exhibit contagion in this task, participants first needed to learn observees' risk-preference from their choices. The proportion of correct predictions in Predict trials and postexperiment ratings about the observees' risk-preference showed that participants indeed succeeded in learning about their risk-preference (Fig. 3A and B). Furthermore, a closer examination of the prediction performance revealed that the learning was faster when the participants' preference was congruent with the observee's (Fig. S6), suggesting that the participants used their own preference as a starting point.

This suggestion is also supported by a formal model-based analysis. We constructed a family of computational models and fit those models to the participants' prediction behavior (SI Text 5 and



Table S2). The model comparison revealed that a Bayesian learning model using each individual's own risk-preference as a prior belief provided a better fit to the participants' actual predictions than alternative models did.

**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 observee'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).

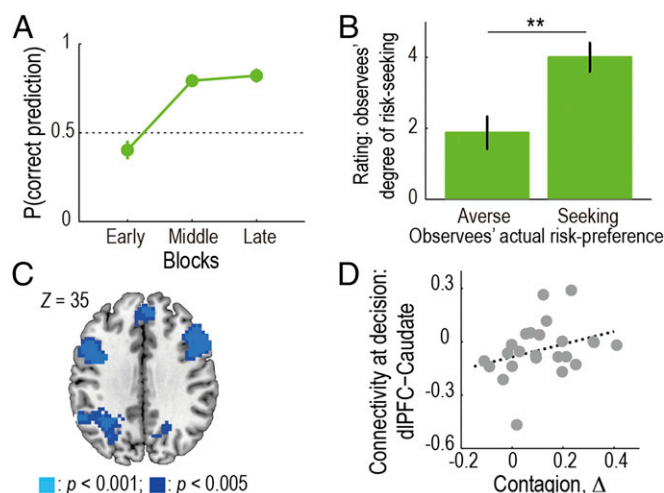
**Functional Connectivity Between the dlPFC and the Caudate.** Given the above results together with previous findings of the anatomical and functional connectivity between dlPFC and caudate (31, 32), we further reasoned that functional connectivity between the two regions would account for individual 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



**Fig. 3.** Learning about others' risk-preferences. (A) Participants' learning curve of the observees' risk-preferences. The proportion of correct predictions in Predict trials is plotted as a function of time (mean  $\pm$  SEM across participants;  $n = 24$ ). Data from Sessions 2 and 4 are collapsed. (B) Post-experiment ratings about the observees' risk-preferences. We plot participants' ratings (six-point scale) about how likely the observee would accept a gamble, separately for risk-averse and -seeking observees (mean  $\pm$  SEM across participants). \*\* $P < 0.01$ . (C) Neural representation of belief-updating during learning about others' risk-preferences. Activity in the dlPFC, the dmPFC, and the IPL significantly correlated with the update signal,  $D_{KL}$ , which is measure of the difference between the posterior and the prior, at the time of confirmation in Observe trials [ $P < 0.05$  FWE corrected at cluster-level; GLM V (SI Methods)]. (D) Functional connectivity between dlPFC and caudate at the time of decision on Self trials reflected the degree of behavioral contagion across participants ( $\rho = 0.38$ ;  $P = 0.035$ ).

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 observees. 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 “anticipation 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 dlPFC. 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 dlPFC 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 dlPFC. The role of a dorsolateral prefrontal–striatal circuit (31) in contagion of risk-preferences can be interpreted within a broader literature implicating this circuit in cognitive functions more generally such as in response-selection, planning, and set-shifting (32). The dlPFC 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. S1A), 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 observee. 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 observees 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 observee’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.).

1. Knutson B, Bossaerts P (2007) Neural antecedents of financial decisions. *J Neurosci* 27(31):8174–8177.
2. Platt ML, Huettel SA (2008) Risky business: The neuroeconomics of decision making under uncertainty. *Nat Neurosci* 11(4):398–403.

3. Bossaerts P, Plott C (2004) Basic principles of asset pricing theory: Evidence from large-scale experimental financial markets. *Rev Finance* 8(2):135–169.
4. Lejuez CW, et al. (2002) Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *J Exp Psychol Appl* 8(2):75–84.



5. Ligneul R, Sescousse G, Barbalat G, Domenech P, Dreher J-C (2013) Shifted risk preferences in pathological gambling. *Psychol Med* 43(5):1059–1068.
6. De Martino B, Kumaran D, Seymour B, Dolan RJ (2006) Frames, biases, and rational decision-making in the human brain. *Science* 313(5787):684–687.
7. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–292.
8. Callen M, Isaqzadeh M, Long JD, Sprenger C (2014) Violence and risk preference: Experimental evidence from Afghanistan. *Am Econ Rev* 104(1):123–148.
9. Cameron L, Shah M (2013) *Risk-Taking Behavior in the Wake of Natural Disasters* (National Bureau of Economic Research, Cambridge, MA).
10. Xue G, Lu Z, Levin IP, Bechara A (2011) An fMRI study of risk-taking following wins and losses: Implications for the gambler's fallacy. *Hum Brain Mapp* 32(2):271–281.
11. Cialdini RB, Goldstein NJ (2004) Social influence: Compliance and conformity. *Annu Rev Psychol* 55:591–621.
12. Albert D, Chein J, Steinberg L (2013) Peer influences on adolescent decision making. *Curr Dir Psychol Sci* 22(2):114–120.
13. Bondt WFM, Thaler R (1985) Does the stock market overreact? *J Finance* 40(3):793–805.
14. Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. *Neuron* 61(1):140–151.
15. Campbell-Meiklejohn DK, Bach DR, Roepstorff A, Dolan RJ, Frith CD (2010) How the opinion of others affects our valuation of objects. *Curr Biol* 20(13):1165–1170.
16. Chung D, Christopoulos GI, King-Casas B, Ball SB, Chiu PH (2015) Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nat Neurosci* 18(6):912–916.
17. Zaki J, Schirmer J, Mitchell JP (2011) Social influence modulates the neural computation of value. *Psychol Sci* 22(7):894–900.
18. Charpentier CJ, Moutsiana C, Garrett N, Sharot T (2014) The brain's temporal dynamics from a collective decision to individual action. *J Neurosci* 34(17):5816–5823.
19. Garvert MM, Moutoussis M, Kurth-Nelson Z, Behrens TEJ, Dolan RJ (2015) Learning-induced plasticity in medial prefrontal cortex predicts preference malleability. *Neuron* 85(2):418–428.
20. Izuma K, Adolphs R (2013) Social manipulation of preference in the human brain. *Neuron* 78(3):563–573.
21. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W (2009) Risk-dependent reward value signal in human prefrontal cortex. *Proc Natl Acad Sci USA* 106(17):7185–7190.
22. Hsu M, Krajbich I, Zhao C, Camerer CF (2009) Neural response to reward anticipation under risk is nonlinear in probabilities. *J Neurosci* 29(7):2231–2237.
23. Tobler PN, Christopoulos GI, O'Doherty JP, Dolan RJ, Schultz W (2008) Neuronal distortions of reward probability without choice. *J Neurosci* 28(45):11703–11711.
24. Rutledge RB, Skandali N, Dayan P, Dolan RJ (2015) Dopaminergic modulation of decision making and subjective well-being. *J Neurosci* 35(27):9811–9822.
25. Tzourio-Mazoyer N, et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273–289.
26. O'Reilly JX, et al. (2013) Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc Natl Acad Sci USA* 110(38):E3660–E3669.
27. Suzuki S, et al. (2012) Learning to simulate others' decisions. *Neuron* 74(6):1125–1137.
28. Boorman ED, O'Doherty JP, Adolphs R, Rangel A (2013) The behavioral and neural mechanisms underlying the tracking of expertise. *Neuron* 80(6):1558–1571.
29. Burke CJ, Tobler PN, Baddeley M, Schultz W (2010) Neural mechanisms of observational learning. *Proc Natl Acad Sci USA* 107(32):14431–14436.
30. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008) Associative learning of social value. *Nature* 456(7219):245–249.
31. Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9(1):357–381.
32. Tekin S, Cummings JL (2002) Frontal-subcortical neuronal circuits and clinical neuropsychiatry: An update. *J Psychosom Res* 53(2):647–654.
33. Bikhchandani S, Sharma S (2000) Herd behavior in financial markets. *Staff Pap Int Monet Fund* 47(3):279–310.
34. Chaboud AP, Chiquoine B, Hjalmarsson E, Vega C (2014) Rise of the machines: Algorithmic trading in the foreign exchange market. *J Finance* 69(5):2045–2084.
35. Mohr PNC, Biele G, Heekeren HR (2010) Neural processing of risk. *J Neurosci* 30(19):6613–6619.
36. Wu CC, Sacchet MD, Knutson B (2012) Toward an affective neuroscience account of financial risk taking. *Front Neurosci* 6:159.
37. Christopoulos GI, Tobler PN, Bossaerts P, Dolan RJ, Schultz W (2009) Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *J Neurosci* 29(40):12574–12583.
38. Li J, Delgado MR, Phelps EA (2011) How instructed knowledge modulates the neural systems of reward learning. *Proc Natl Acad Sci USA* 108(1):55–60.
39. Fitzgerald THB, Seymour B, Bach DR, Dolan RJ (2010) Differentiable neural substrates for learned and described value and risk. *Curr Biol* 20(20):1823–1829.
40. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF (2005) Neural systems responding to degrees of uncertainty in human decision-making. *Science* 310(5754):1680–1683.
41. Weber BJ, Huettel SA (2008) The neural substrates of probabilistic and intertemporal decision making. *Brain Res* 1234:104–115.
42. Symmonds M, Wright ND, Bach DR, Dolan RJ (2011) Deconstructing risk: Separable encoding of variance and skewness in the brain. *Neuroimage* 58(4):1139–1149.
43. Labudda K, et al. (2008) Neural correlates of decision making with explicit information about probabilities and incentives in elderly healthy subjects. *Exp Brain Res* 187(4):641–650.
44. O'Doherty J, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304(5669):452–454.
45. Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. *J Neurosci* 27(31):8161–8165.
46. Pizzagalli DA, et al. (2009) Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *Am J Psychiatry* 166(6):702–710.
47. Maner JK, et al. (2007) Dispositional anxiety and risk-avoidant decision-making. *Pers Individ Dif* 42(4):665–675.
48. Giorgetta C, et al. (2012) Reduced risk-taking behavior as a trait feature of anxiety. *Emotion* 12(6):1373–1383.
49. Bursztyn L, Ederer F, Ferman B, Yuchtman N (2014) Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica* 82(4):1273–1301.
50. Burke CJ, Tobler PN, Schultz W, Baddeley M (2010) Striatal BOLD response reflects the impact of herd information on financial decisions. *Front Hum Neurosci* 4:48.
51. Tamir DI, Mitchell JP (2010) Neural correlates of anchoring-and-adjustment during mentalizing. *Proc Natl Acad Sci USA* 107(24):10827–10832.
52. Buckner RL, Carroll DC (2007) Self-projection and the brain. *Trends Cogn Sci* 11(2):49–57.
53. Mitchell JP (2009) Inferences about mental states. *Philos Trans R Soc Lond B Biol Sci* 364(1521):1309–1316.
54. McNamee D, Liljeblom M, Zika O, O'Doherty JP (2015) Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: A multivariate fMRI study. *J Neurosci* 35(9):3764–3771.
55. Tanaka SC, Balleine BW, O'Doherty JP (2008) Calculating consequences: Brain systems that encode the causal effects of actions. *J Neurosci* 28(26):6750–6755.
56. Morris RW, Dezfouli A, Griffiths KR, Balleine BW (2014) Action-value comparisons in the dorsolateral prefrontal cortex control choice between goal-directed actions. *Nat Commun* 5:4390.
57. Chari VV, Kehoe PJ (2004) Financial crises as herds: Overturning the critiques. *J Econ Theory* 119(1):128–150.
58. Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J Polit Econ* 100(5):992–1026.
59. Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19(2 Pt 1):430–441.
60. Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29(3):162–173.
61. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46(4):1004–1017.
62. Daw ND (2011) Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning Attention and Performance XXIII*, eds Delgado MR, Phelps EA, Robbins TW (Oxford Univ Press, Oxford), pp 3–38.
63. Efron B (1992) Jackknife-after-bootstrap standard errors and influence functions. *J R Stat Soc Ser B Stat Methodol* 54:83–127.
64. Martin LN, Delgado MR (2011) The influence of emotion regulation on decision-making under risk. *J Cogn Neurosci* 23(9):2569–2581.
65. Berg J, Dickhaut J, McCabe K (2005) Risk preference instability across institutions: A dilemma. *Proc Natl Acad Sci USA* 102(11):4209–4214.
66. Kolling N, Wittmann M, Rushworth MFS (2014) Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* 81(5):1190–1202.
67. Cohn A, Engelmann J, Fehr E, Maréchal MA (2014) Evidence for countercyclical risk aversion: An experiment with financial professionals. *Am Econ Rev* 105(2):860–885.
68. Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB (2003) Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage* 19(4):1439–1448.
69. Huettel SA, Stowe CJ, Gordon EM, Warner BT, Platt ML (2006) Neural signatures of economic preferences for risk and ambiguity. *Neuron* 49(5):765–775.
70. van Leijenhorst L, Crone EA, Bunge SA (2006) Neural correlates of developmental differences in risk estimation and feedback processing. *Neuropsychologia* 44(11):2158–2170.
71. Lee TMC, Leung AWS, Fox PT, Gao J-H, Chan CCH (2008) Age-related differences in neural activities during risk taking as revealed by functional MRI. *Soc Cogn Affect Neurosci* 3(1):7–15.
72. Engelmann JB, Tamir D (2009) Individual differences in risk preference predict neural responses during financial decision-making. *Brain Res* 1290:28–51.
73. Smith BW, et al. (2009) Neural substrates of reward magnitude, probability, and risk during a wheel of fortune decision-making task. *Neuroimage* 44(2):600–609.
74. Xue G, et al. (2009) Functional dissociations of risk and reward processing in the medial prefrontal cortex. *Cereb Cortex* 19(5):1019–1027.
75. Paulus MP, et al. (2001) Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty. *Neuroimage* 13(1):91–100.
76. Volz KG, Schubotz RI, von Cramon DY (2003) Predicting events of varying probability: Uncertainty investigated by fMRI. *Neuroimage* 19(2 Pt 1):271–280.
77. Matthews SC, Simmons AN, Lane SD, Paulus MP (2004) Selective activation of the nucleus accumbens during risk-taking decision making. *Neuroreport* 15(13):2123–2127.
78. Volz KG, Schubotz RI, von Cramon DY (2004) Why am I unsure? Internal and external attributions of uncertainty dissociated by fMRI. *Neuroimage* 21(3):848–857.
79. Cohen MX, Heller AS, Ranganath C (2005) Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain Res Cogn Brain Res* 23(1):61–70.
80. Huettel SA, Song AW, McCarthy G (2005) Decisions under uncertainty: Probabilistic context influences activation of prefrontal and parietal cortices. *J Neurosci* 25(13):3304–3311.
81. Huettel SA (2006) Behavioral, but not reward, risk modulates activation of prefrontal, parietal, and insular cortices. *Cogn Affect Behav Neurosci* 6(2):141–151.
82. Mohr PNC, Biele G, Krugel LK, Li S-C, Heekeren HR (2010) Neural foundations of risk-return trade-off in investment decisions. *Neuroimage* 49(3):2556–2563.
83. Bach DR, Hulme O, Penny WD, Dolan RJ (2011) The known unknowns: Neural representation of second-order uncertainty, and ambiguity. *J Neurosci* 31(13):4811–4820.
84. Payzan-LeNestour E, Dunne S, Bossaerts P, O'Doherty JP (2013) The neural representation of unexpected uncertainty during value-based decision making. *Neuron* 79(1):191–201.