J Risk Uncertain (2014) 49:125–140
DOI 10.1007/s11166-014-9199-7

The neural correlates of contractual risk and penalty framing

W. Gavin Ekins · Andrew M. Brooks · Gregory S. Berns

Published online: 13 November 2014
© Springer Science+Business Media New York 2014

Abstract Standard economic theory treats contractual risk the same as risk experienced in a lottery, but the transfer of risk from principal to agent may change the perception of risk. Previous experimental studies have shown that positive and negative framing affects both gambles and incentive contracts. An agent's perception of bonus and penalty framing in a contract can determine the extent to which lottery risk and contractual risk are similar. We designed an experiment that tested the effect of bonus and penalty framed contracts on behavior under an implicit chance of failure. Moreover, we used functional magnetic resonance imaging (fMRI) to observe brain activity while participants viewed the contracts and purchased precautionary measures. We found that the dorsal striatum was more active during a penalty frame than a bonus frame. The study suggests that risk experienced by agents in an incentive contract is not comparable to risk experienced as a lottery.

Keywords Risk · Incentive contracts · Framing effects · fMRI · Striatum

JEL Classifications C91 · D81 · D86 · D87

This publication was made possible by grant number R01 DA0250445 from the National Institute of Drug Abuse at the National Institutes of Health. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Drug Abuse.

W. G. Ekins (✉)
Department of Economics and Center for Neuropolicy, Emory University, 1602 Fishburne Dr, Atlanta, GA 30322, USA
e-mail: wekins@marymount.edu

A. M. Brooks
Center for Neurpolicy, Emory University, 36 Eagle Row, Atlanta, GA 303022, USA
e-mail: abrook6@emory.edu

G. S. Berns
Department of Economics and Center for Neuropolicy, Emory University, 1602 Fishburne Dr, Atlanta, GA 30322, USA
e-mail: gberns@emory.edu

The search for the neural substrates that underlie risky decision making has been a centerpiece of neuroeconomic research. Lotteries have largely been the vehicle by which risk preferences are studied, but the discrete choice between lotteries does not reflect the entire spectrum of risk that exists in the economy. Risk is often shared between economic entities through contractual obligations (Allen and Lueck 1995). These obligations are designed to shift risk to the agent with the greatest ability to mitigate risk, but in the process, the contract may change the perception of the risks involved. As such, agents with the ability to mitigate risk may under or overestimate the risks due to the design of the contract. How these agents perceive risk from their contractual obligations may be fundamentally different from the risk experience in a choice between lotteries.

Incentive contracts are one method of shifting risk from a principal to an agent. In practice, the compensation of incentive contracts is framed as bonuses, penalties, or a combination of both (Lewis 1980). Bonus contracts are structured such that the agent receives additional compensation if the terms of the contract have been completed. Conversely, penalty contracts disburse compensation upfront, but if the agent fails to perform, then he is required to pay a fine. Both bonus and penalty framed incentive contracts shift risk to the agent by imposing a cost for failing to complete the contract. In fact, bonus contracts can be reframed as a penalty and vice versa without changing the risk imposed on the agent (Goetz and Scott 1977). For example, if a bonus contract pays an agent $100 upfront plus $50 when the terms of the contract are completed, then it is possible to transform this contract into a penalty contract by paying the agent $150 upfront and fining him $50 if he fails to complete the terms. In either incentive contract, the contractor receives $150 for successfully completing the contract while he receives only $100 if he fails. Therefore, the contractor risks $50 if he fails, regardless of the contractual frame.

Despite their economic equivalents, bonus framing tends to be used more often than penalty framing in labor contracts (Lazear 1991). The prevalence of these contracts may stem from labor's preference for bonus frames. Several studies have shown that individuals prefer contracts structured as a bonus over those structured as a penalty (Luft 1994; Frederickson and Waller 2005). Behavioral experiments of bonus and penalty contracts offer a possible explanation for these preferences. In experiments where participants use a monetary proxy for effort, which reduced the probability of failing to complete their contract, participants exerted more effort under a penalty contract than a bonus contract (Hannan et al. 2005; Brink 2008).

Brink (2008) has argued that the difference between contractual frames is a function of loss aversion, but the ability of agents to control their risk through effort, also known as precaution in the law and economics literature, may fundamentally change the perception of risk. If agents view an increase in effort as simply a choice between lotteries with different probabilities, then contractual framing should mirror framing effects in gain and loss gambles. In both cases, framing changes the reference point on the agent's preference curve (Luft 1994), which places the agent into the negative realm and exposes her to loss aversion. Alternatively, if agents view the risk as a function of ability, then overconfidence may be responsible for the observed differences between penalty and bonus contracts (Gervais et al. 2011). How an individual perceives and processes the difference between bonus and penalty contracts may determine the extent to which contractual risk is similar to risk perceived in a lottery.

From the current literature, it is not clear if the behavioral difference between bonus and penalty contracts is a function of loss aversion or overconfidence. Both loss aversion and overconfidence may display similar behaviors. The brain offers a possible method for distinguishing which process is responsible. The circuits in the brain that relate to loss aversion have been well studied in neuroscience. If loss aversion is responsible for the behavioral differences between bonus and penalty contracts, then similar circuits should be active when a participant makes a similar decision to the experiment in Brink (2008). If the decision is related to another circuit, then an alternative explanation, such as overconfidence, may be responsible for behavioral differences.

Neuroscientists have argued that loss aversion is driven by biological factors, or at least facilitated by neurobiology. Tom et al. (2007) found greater activity in the ventral striatum when participants were faced with a possible loss in a gamble compared to a gain. Other researchers have found greater activity in the amygdala and insula during a decision between losses (Yacubian et al. 2006) while the ventral striatum had greater activity during a decision between gains. Furthermore, in studies of participants with brain lesions in the amygdala, researchers found no discernible loss aversion among their participants (Martino et al. 2010). This suggests that the amygdala, insula, and ventral striatum have a role in encoding value and may be the underlying cause of loss aversion. If increased activity in the amygdala, insula, and ventral striatum are correlated with behavioral difference between bonus and penalty contracts, we can conclude that loss aversion is likely the culprit.

The dorsal striatum has also been shown to encode value, but it has largely been connected to evaluating goal-directed action or action-contingent learning (Balleine et al. 2007). The dorsal striatum has been shown to encode reward prediction error during goal-directed tasks in both humans and primates (Delgado et al. 2003; Delgado 2007). In particular, O'Doherty et al. (2004) argued that the ventral striatum encodes value of an outcome while the dorsal striatum encodes the value of an action. If the dorsal striatum is the region where activity differs between bonus and penalty contracts, then it is possible that the difference in behavior between the two contracts is a prediction error of the required action. Thus, overconfidence may be responsible.

In this study, we explore the effect of contractual framing and the cost of failure on participants' perception of risk. Moreover, we used functional magnetic resonance imaging (fMRI) to explore the regions of the brain involved in the valuation of contractual obligations. In the following sections, we developed an experiment, based on incentive contracts, to study these behavioral asymmetries without explicitly inducing the probability of failure. Instead, we used a task that relies on the participants' assessment of their ability and a costly effort, which does not explicitly reduce the probability of failure. We used fMRI data, acquired during the experiment, to track the cognitive circuits that were active during decisions over different contractual frames.

1 Experimental design

1.1 Procedure

We recruited 30 healthy, right-handed participants from Emory University (M: 15, F: 15, age: 18–41). Two of these participants had to be removed from the data because of

excessive head movement during image acquisition. All procedures were approved by the Emory University IRB, and each participant provided written consent and was asked to complete three surveys: Risk Preference Worksheet, EPQ-R, and BIS/BAS.[1] All instructions were presented on a computer terminal with periodic quizzes to ensure participant's comprehension of the experiment. In the experiment, participants were contracted to answer trivia questions correctly. Trivia questions from the Who Wants to Be a Millionaire board game were displayed as a multiple choice quiz with 4 possible answers. The terms of the contract required the participants to answer 10 out of 10 trivia questions correctly. Before participants were asked to perform the contract, i.e. answer 10 questions correctly, they were given the opportunity to purchase additional questions. Each additional question allowed the participant to answer a question incorrectly without failing to complete the terms of the contract.

We manipulated the framing of the contract as well as the cost of failing to complete the contract. The 2x3 factorial design captured the effects of the contract framing and the cost of failure as well as any interaction effects (Table 1). In the first factor, we framed the contract as either a penalty or a bonus. In the second factor, we varied the cost of failure: high, medium, and low. The benefit of completing the contract was consistently $30 for all contracts, but the cost of failure dictated the amount a participant received if he failed to complete the contract: $45, $30 and $15. If the cost of failure was $15 and the participant failed to complete the contract, then he received $15 from the experimenter. Likewise, if the cost of failure was $45 and the participant failed to complete the contract, then he had to pay the experimenter $15 from his endowment.

In the bonus treatments, participants received either -$15, $0, or $15 upfront[2] depending on the cost of failure treatment and an additional $45, $30, or $15, respectively, if they were able to complete the contract. Thus, there was no cost to the participant for failing in the bonus treatments. In this case, if the participant failed to complete the contract, they incurred the opportunity cost of losing the bonus. In the penalty treatments, participants received $30 upfront and no additional payment if they were able to complete the contract. If the participant failed, he had to return $15, $30, or $45 to the experimenter based on the cost of failure treatment. Regardless of the frame, participants received $30 if the contract was completed and -$15, $0, or $15 if they failed, depending on the cost of failure treatment. As such, after the endowment is adjusted for the upfront payment, all bonus treatments displayed a gain for completing the contract ($15, $30, $45) while all penalty treatments displayed a loss for failing to complete the contract (-$15, -$30, -$45).

The experiment was separated into three sections. In the first section of the experiment, participants answered trivia questions to earn an endowment and to gauge their ability to answer the trivia questions correctly. In the second section, participants were shown different incentive contracts, based on the treatments, after which they were given the chance to purchase additional trivia questions. In the third section of the experiment, participants were randomly assigned a contract

[1] Eysenck Personality Questionnaire—1985 revision (EPQ-R), Behavioral Inhibition System / Behavioral Activation System (BIS/BAS) test.

[2] The negative sign in the upfront amount represents the amount surrendered to the experimenter. This represents a type of escrow.

Table 1 Experiment parameters by treatments

Factor 1	Factor 2	Upfront	If successful	If failed	Total successful	Total failed	Cost of failure
Bonus	Low	15	15	0	30	15	15
	Medium	0	30	0	30	0	30
	High	−15	45	0	30	−15	45
Penalty	Low	30	0	−15	30	15	15
	Medium	30	0	−30	30	0	30
	High	30	0	−45	30	−15	45

to perform. Participants had to answer the appropriate number of questions correctly.

In the first section, participants earned $30 by correctly answering trivia questions from the Who Wants to Be a Millionaire board game. Each question answered correctly earned the participant $2. Difficulty of trivia questions were titrated such that the participants were correct approximately 50 % of the time (Table 2). There were 10 levels of question difficulty. A moving-average adjustment algorithm was employed to assign the appropriate level of difficulty to each participant. The number of correct answers over the total number of questions answered was displayed at the top of the screen to provide the participants with a rough estimate of their probability of answering a trivia question correctly. The initial phase ended when the participant had earned $30, i.e. the participant correctly answered 15 trivia questions correctly.

In the second section of the experiment, the participants were shown 60 contracts, 10 repetitions of the 6 treatments. After each contract was presented, participants were given the opportunity to purchase additional trivia questions at a cost of $0.50 each. During this section of the experiment, fMRI images were acquired while participants viewed the contracts and decided the amount of additional questions to purchase. Each contract consisted of four screens: fixation, passive contract, decision, and review; in order of display (Figs. 1 and 2). The participants were required to view the passive contract phase for a minimum of 4 sec after which they used a button to move forward. During the decision phase, participants were able to purchase additional trivia questions. The number of additional questions purchased along with the total cost was displayed in the lower half of the screen while the contract terms were visible in the upper half. This phase was not timed and participants chose to move forward by pressing a button. During the review screen, participants reviewed their decision for 2 sec. During the fixation screen, participants viewed a blank screen for 2–6 sec, uniformly distributed. After the fixation screen, the next trial began with a new contract.

After the completion of the second section of the experiment, the participants rolled dice to determine which of the contracts they performed. Participants performed the contract with the number of additional questions chosen in the second section of the experiment. When the participant no longer had a sufficient number of questions to complete the contract or the participant had answered the 10 questions correctly, the participant was paid according to the terms of the contract minus the cost of the additional questions purchased.

Table 2 Average probability of answering a trivia question correctly by participant

Participant	Questions answered incorrectly	Probability of correct answer	Contract completed
1	47	24.20 %	Completed
2	20	42.90 %	Completed
3	20	42.90 %	Completed
4	18	45.50 %	Failed
5	19	44.10 %	Failed
6	11	57.70 %	Completed
7	20	42.90 %	Completed
8	18	45.50 %	Failed
9	23	39.50 %	Completed
10	21	41.70 %	Completed
11	19	44.10 %	Failed
12	17	46.90 %	Failed
13	18	45.50 %	Failed
14	14	51.70 %	Completed
15	11	57.70 %	Completed
16	14	51.70 %	Completed
17	13	53.60 %	Completed
18	13	53.60 %	Completed
19	13	53.60 %	Completed
20	15	50.00 %	Failed
21	18	45.50 %	Failed
22	11	57.70 %	Completed
23	12	55.60 %	Completed
24	12	55.60 %	Completed
25	13	53.60 %	Completed
26	15	50.00 %	Failed
27	20	42.90 %	Completed
28	15	50.00 %	Failed
29	25	37.50 %	Completed
30	18	45.50 %	Failed
Mean	17.43	47.60 %	63 %
St. Dev.	6.72	7.20 %	

1.2 Behavioral model

To calculate the theoretically optimal amount of additional trivia questions to purchase, the total number of questions in the first section of the experiment for each participant was used as a proxy for the perceived probability of answering a question correctly. We calculated the probability of completing the contract with the binomial model (1), where a was the total additional trivia questions purchased, q was the required number of questions to answer, b was the required endowment questions (first at 15 for this experiment), and t was the total number of trivia questions answered in the initial

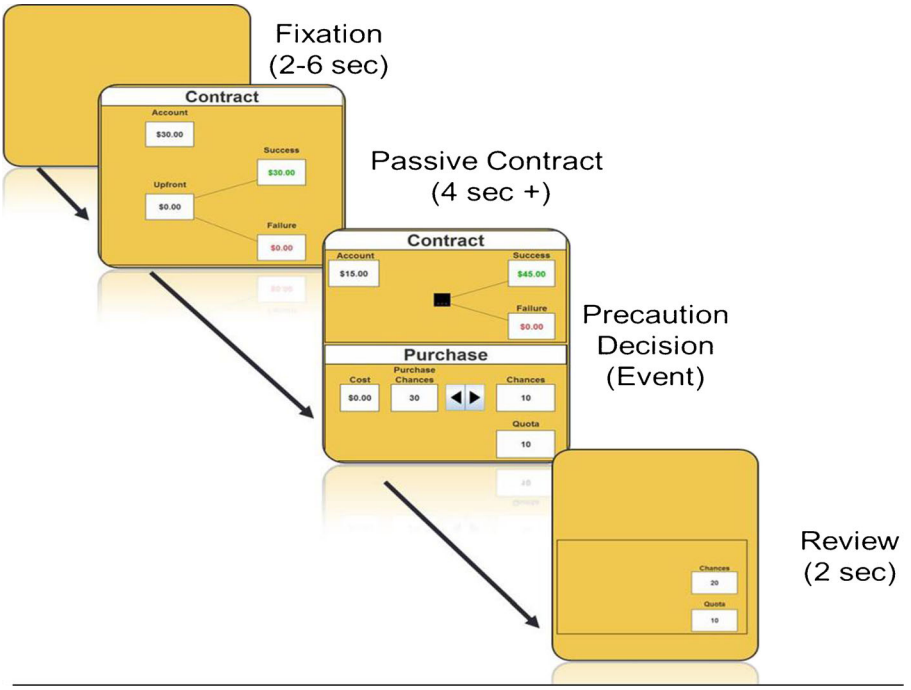


Fig. 1 Experiment software screenshots and trial timing

portion of the experiment. Note that b/t is an estimate for the perceived probability of incorrectly answering a trivia question. We calculated the optimal number of additional trivia questions to purchase by using a risk neutral, expected utility model (2), where *CF* is the outcome if the contract is incomplete. The optimal number of additional questions (a^*) was calculated using iterative methods (Table 3).

$$\rho(a,t) = \binom{q+a}{q}\left(\frac{b}{t}\right)^{q}\left(1-\frac{b}{t}\right)^{a} \tag{1}$$

$$\pi_s(OC) = \max_a R\rho(a,t_s) + CF(1-\rho(a,t)) - ca \tag{2}$$

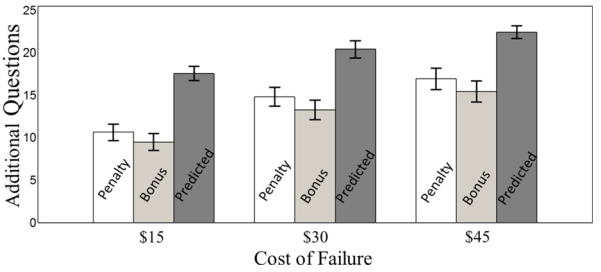


Fig. 2 Additional questions purchased by treatment

Table 3 Additional questions purchase by treatment

	All			Bonus			Penalty		
Purchased	13.51			12.74			14.28		
	(6.500)			(6.443)			(6.505)		
Opportunity cost	$15	$30	$45	$15	$30	$45	$15	$30	$45
Purchased	10.06	14.14	16.33	9.50	13.27	15.44	10.62	14.81	16.93
	(5.109)	(6.060)	(6.702)	(5.158)	(6.157)	(6.676)	(5.100)	(5.955)	(6.740)
Predicted	17.57	20.39	22.43						
	(4.384)	(5.259)	(3.985)						

Standard Deviation in Parentheses

1.3 fMRI Data acquisition and analysis

We used blood-oxygen level dependent functional magnetic resonance imaging (BOLD-fMRI) to measure neural activity of participants while performing our task. BOLD-fMRI measures changes in local magnetic fields due to the presence of oxygen in the blood. When neurons fire in the brain, oxygen-rich blood flows to the active region to replenish the neurons. The change in oxygenated blood affects the local magnetic field in a manner such that the intensity of the MRI signal changes in proportion to the local concentration of oxygenated-hemoglobin. Moreover, this signal change follows a well-defined pattern called the canonical hemodynamic response function (HRF), which we denote as h(t). By mapping signals from the BOLD response onto the HRF, we can determine the approximate time and level of activity within a voxel, a unit of 3-dimensional space similar to a pixel.

The properties of the BOLD signal are such that multiple sources of neural activity can be linearly combined (Boynton et al. 1996). In other words, the BOLD signal can be viewed as a sum of multiple HRFs stemming from activations at a particular time and location, which we denote as *a(t)*. The level of BOLD response related to a stimulus at time period t is approximated by a convolution of the HRF of all stimuli (3).

$$b(t) = \int_0^\infty h(u)a(t-u)du \tag{3}$$

Functional imaging was performed with a Siemens 3-Tesla Trio whole-body scanner. We obtained two types of images. The first image, known as a T1-weighted structural image, was a high-resolution image of the participants' brain structures. These images[3] were acquired for each participant prior to the second portion of the experiment. The second image, known as a T2*-weighted echo-planar image, registered the signal change due to the BOLD response. These images[4] were acquired while the participants were viewing the contracts and purchasing additional questions.

[3] TR =2,600 ms, TE =3.02 ms, flip angle =8°, 240×256 matrix, 176 sagittal slices, 1 mm^3 voxel size.
[4] TR =2,000 ms, TE =30 ms, flip angle =73°, FOV =192 mm×192 mm, 64×64 matrix, 33 3.5-mm thick axial slices, and 3×3×3.5 mm voxels.

The BOLD images were subjected to standard preprocessing commonly used in neuroimaging. First, we corrected for the participant's head motion within the scanner. Second, we temporally aligned the images, known as slice timing correction, which accounts for the different times from which each location in the brain is measured. Third, we spatially registered the images using landmarks in the brain so that we could compare images between participants. This process normalized the images to the Montreal Neurological Institute (MNI) template brain. Finally, the images were smoothed to improve signal to noise ratio of individual voxels, which allows for slight misalignments between participants.[5]

Although we cannot observe individual neurons firing, we can search for clusters of neurons that were active during a particular stimulus. In other words, we sought a collection of spatially contiguous voxels with similar BOLD responses stemming from a particular event. In order to isolate the BOLD response, we controlled for events unrelated to the event of interest. Let the BOLD signal from the HRF of event i be denoted as $S_i(t)$ .In addition, let $c_i(t)$ be additional controls for head movement and time trends. Given the linearity of the BOLD signal, we can estimate the level of activation from an event by regressing the HRF of all events with controls on the BOLD signal for each voxel (v). Our BOLD signal regression model was structured as the sum of both HRF, controls, and a constant (4). The error term of (4) followed an AR(1) process.

$$b_v(t) = \propto + \sum_i \beta_i^v S_i(t) + \sum_i \gamma_i^v c_i(t) + \varepsilon(t) \tag{4}$$

In our experiment, we have 4 event types for which we tracked the hemodynamic response in our model (5). In each period, the participant viewed the contract passively (*PC*), decided the amount of additional questions to purchase (*BD & PD*), and reviewed their decision (*R*). We also split the passive contract phase and the active decision phase events into separate events depending on the framing of the contract, bonus or penalty. Thus, we were able to compare the neural activity between a bonus and penalty contract during the passive and active phases of the experiment. Initially we separated the passive phase into both penalty and bonus frames, but the lack of significant difference between the passive bonus and the passive penalty phase and the increase in error led us to combine the passive phase into one event. In addition, we developed a model, (6), to account for the cost of failure (*CF*) treatments. In this model, the decision phase of the experiment is split by framing and by cost of failure. This model included the framing of bonus and penalty and the three levels of the cost of failure: 15, 30, and 45.

$$b_v(t) = \propto + \beta_{PC}^v S_{PC}(t) + \beta_{BD}^v S_{BD}(t) + \beta_{PD}^v S_{PD}(t) + \beta_R^v S_R(t) + \sum_i \gamma_i^v c_i(t) + \varepsilon(t) \tag{5}$$

$$b_v(t) = \propto + \beta_{PC}^v S_{PC}(t) + \sum_{i \in CF} \beta_{BD,i}^v S_{BD,i}(t) + \sum_{i \in CF} \beta_{PD,i}^v S_{PD,i}(t) + \sum_i \gamma_i^v c_i(t) + \varepsilon(t) \tag{6}$$

where

$$CF = \{15, 30, 45\}$$

[5] Isotropic Gaussian kernel, full-width half-maximum = 8 mm.

Penny et al. (2007) provided the standard method within neuroscience for estimating the magnitude and significance of our models' coefficients. In this method, each participant's regression coefficients were estimated independently. The coefficients from each participant were then averaged and a *t*-test was used to estimate significance. In this case, participant specific error is absorbed into the constant of each participant's regression model. We used SPM[6] to estimate the coefficients from (6).

Since each regression model does not consider the BOLD signal of the surrounding voxels, significance of any one voxel could be a function of smoothing or random search. Standard methods for correcting multiple comparisons, such as Bonferroni, would increase significance to unobtainable thresholds due to the larger number of voxels in the brain. Given the properties of the hemodynamic response to neural activity, larger clusters were less likely to be a function of smoothing or random search. By setting an appropriate cluster threshold, we can avoid the false detection of elevated neural activity by eliminating sufficiently small clusters. We used Monte Carlo simulations[7] with the smoothness parameters from the neuroimages to estimate a clusters size threshold such that the false detection rate (FDR) was less than 0.05 for a given BOLD signal height threshold.

2 Results

2.1 Behavioral results

The additional questions purchased by participants suggested that both the framing and the cost of failure affected the participants' perceptions of the cost of failure. An increase in the opportunity cost significantly increased the amount of additional questions purchased ($F[2,1799] = 48.24$, $p<0.001$). Similarly, framing the contract as a penalty also increased the preferred number of additional questions ($F[1,1799] = 290.24$, $p<0.001$). The interaction between the two factors did not significantly predict the number of additional questions purchased ($F[2,1799] = 0.96$, $p=0.385$). This suggests that the penalty frame does not change the slope of the additional questions purchased with respect to the cost of failure but linearly shifts the intercept of the curve.

Participants purchased fewer additional questions than the model predicted. Participants would maximize their expected utility by purchasing 17, 20, or 21 additional questions under a cost of 15, 30, or 45, respectively (Table 3). In the bonus treatments, participants purchased 9.64, 13.52, or 15.65 additional questions on average in the 3 opportunity cost treatments, all of which are significantly less than the predicted amount ($t_{10}=-18.37$, $p<0.001$; $t_{10}=-15.97$, $p<0.001$; $t_{10}=-12.66$; $p<0.001$ respectively). In the penalty treatments, participants purchased 10.99, 15.31, and 17.41 for the three costs of failure. Similarly, all of the penalty treatments were significantly less than the predicted amount ($t_{10}=-14.49$, $p<0.001$; $t_{10}=-11.08$, $p<0.001$; $t_{10}=-7.98$; $p<0.001$ respectively).

[6] Statistical Parametric Mapping (SPM) is software commonly used in neuroscience.
[7] AlphaSim, AFNI

Participants followed a similar pattern of diminishing returns for purchasing an additional question as shown in the model. In the model, an increase in the opportunity cost from 15 to 30 increased the optimal number of questions purchased by 3, while an increase from 30 to 45 only increased the optimal number of questions by 1. Similarly, we observed a greater change in the additional questions purchased between the 15 and 30 opportunity cost treatments than the 30 and 45 treatments. In both the penalty and the bonus treatments, there was roughly double the change in additional questions purchased from 15 to 30 as there was in the change between 30 and 45.

Neither the contractual frame nor the cost of failure affected the participants' reaction times. There were no significant differences in reaction time during the passive contract phase for any of the treatments. Moreover, participants rarely viewed the passive screen for more than the minimal amount of time, 4 sec. In the decision phase, the reaction time increased as the cost of failure increased, but this was only marginally significant ($F[2,1799] = 2.570$, $p = 0.077$). There was no interaction effect between the contractual framing and cost of failure on reaction times.

The average amount of additional questions purchased for each participant did not correlate with measures of loss aversion or risk aversion from the surveys. The surveys used hypothetical questions and were not salient. This may explain the lack of correlation. The outcomes of the other two surveys, the BIS/BAS and the EPQ-R, also did not correlate with additional questions purchased.

2.2 Neuroimaging

Using the images obtained while participants were making their decisions, we examined the difference in brain activity when a penalty contract was presented versus that of a bonus contract. In a whole brain analysis, we found 5 clusters with significant differences (height threshold $p<0.005$, cluster size ≥ 79 voxels) in the (BOLD) signal when contrasting the active decision phase of the two framing treatments (Table 4). The most significant differences were in the dorsal striatum (Max T=6.23; [x=−6; y=−1; z=13]). While making decisions in a bonus contract, the percent signal change in the dorsal striatum cluster was not significantly greater than zero (Fig. 3b). Conversely, the percent signal change during the penalty contract was significantly greater than zero,

Table 4 Clusters from whole brain analysis

Structure	L/R/B	Voxels	Peak Voxel (MNI)			Max T
			X	Y	Z	
Penalty>Bonus						
Dorsal striatum	B	230	−6	−1	13	6.23
Posterior insula	L	326	−48	−28	13	4.21
Posterior insula	R	286	36	−19	−15	4.07
Motor cortex	L	192	−24	−37	73	4.62
Motor cortex	R	469	30	−31	76	4.73
Bonus>Penalty						
Visual cortex	B	360	9	−88	20	4.35

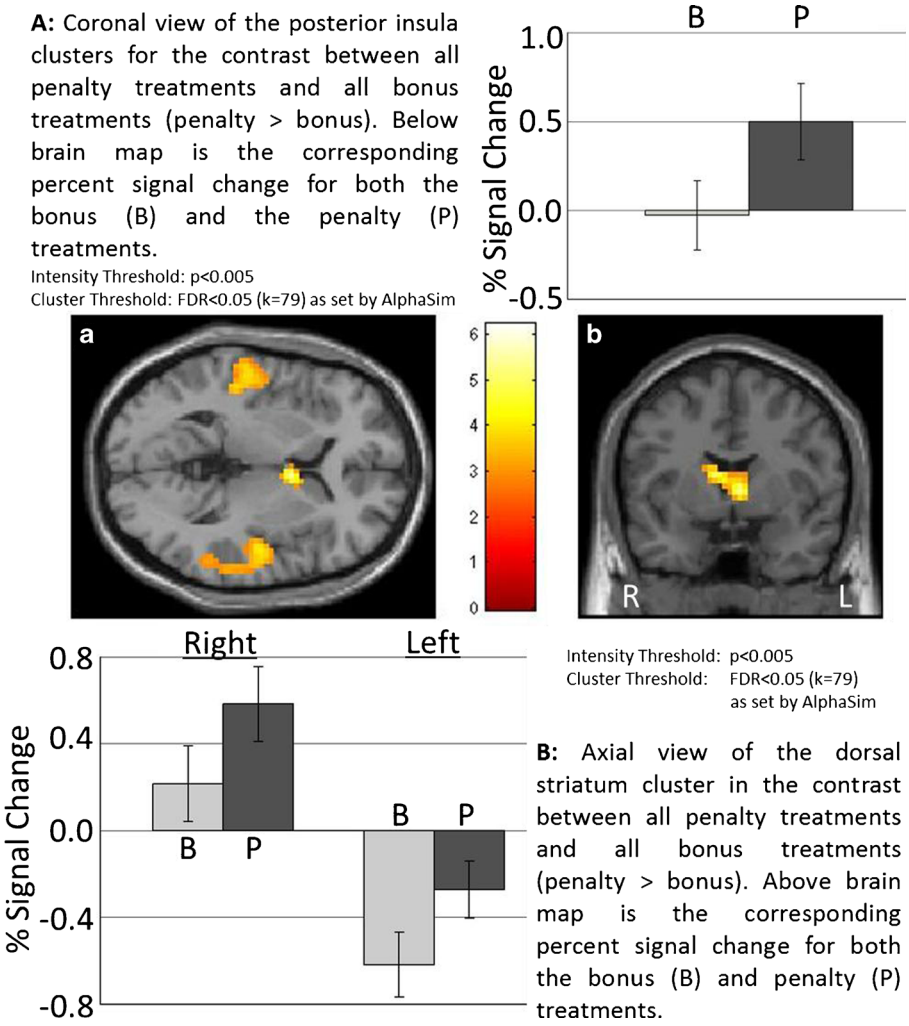


Fig. 3 Brain maps and signal intensity charts

suggesting that this region was active when participants were purchasing additional questions in a penalty contract but not in bonus contracts.

As expected, two clusters were located in the motor cortex. Given that there were more button presses during the penalty than the bonus contract, we expected increased activity in the motor cortex. The remaining two clusters compose bilateral posterior insula (Fig. 3a). The right cluster showed increased BOLD signal during the penalty contract with respect to the bonus contract, but both were significantly greater than zero. In the opposite contrast, we observed a bilateral cluster in the visual cortex.

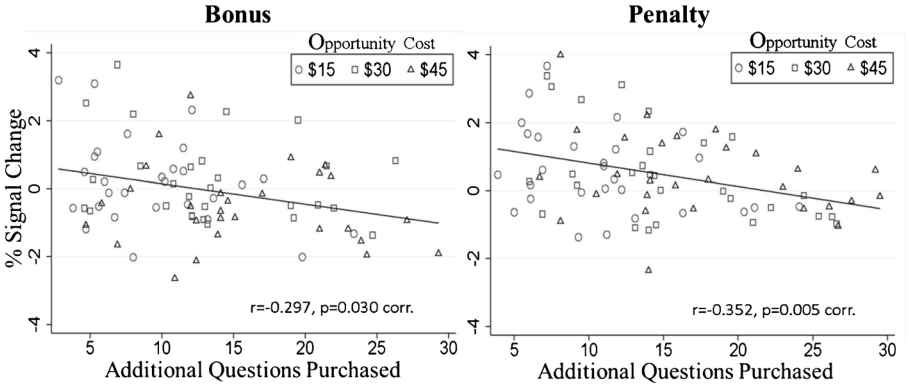
To ensure that the 5 observed clusters were not driven by the difference in the button presses between the 2 treatments, we extracted the average signal change from the 5 clusters to compare with participants' behavior. To control for different incentives under the different opportunity costs of failure, we extracted the average signal change for each of the 6 treatments and compared them with the average additional questions

Table 5 Correlations between brain regions and additional questions purchased

Structure	Bonus Add. questions		Penalty Add. questions	
	r	*p*-value	r	*p*-value
Dorsal Striatum	−0.297	0.030	−0.352	**0.005**
Posterior Insula	−0.165	0.673	−0.207	0.296
Posterior Insula	−0.034	1.000	−0.101	1.000
Motor Cortex	−0.201	0.337	−0.218	0.234
Motor Cortex	0.035	1.000	−0.028	1.000

purchased in each treatment. In other words, the average BOLD signal in the penalty contract with a $15 cost of failure is compared with the average additional questions purchased in the same treatment. Therefore, the bonus and the penalty treatments have 86 data points each (28 participant×3 cost of failure treatments).

Since the participant could increase and decrease the additional questions in the decision phase, the number of button presses was often more than the additional questions purchased. Although the additional questions purchased and button presses were strongly correlated (r=0.774), the discrepancy in button presses and additional questions purchased may help to explain if the differences in brain activity are a function of the decision or actuating the decision. Neither of the motor cortex clusters significantly correlated with the additional questions purchased (Table 5), but the left motor cortex cluster significantly correlated with the number of button presses (r= −0.238, p_{corr}<0.05). The posterior insula clusters also failed to significantly correlate with the additional questions purchased or the button presses. The dorsal striatum, in both the bonus treatments and the penalty treatments, negatively correlated with the additional questions purchased (Fig. 4), even when controlling for multiple comparisons. Although the dorsal striatum also correlated with the button presses, it was weaker than the additional questions purchased and more likely due to the correlation



Scatter plots of average additional questions purchased in each treatment vs. percent signal change in the dorsal striatum cluster for each treatment regressor. In both the bonus and penalty treatments, the number of questions purchased correlates with the signal change in the dorsal striatum. There is no significant difference between the correlations in the bonus and the penalty treatments [F(1,164)=0.08, p=0.782].

Fig. 4 % Signal change in dorsal striatum custer vs. behavior

between button presses and additional questions purchased. The correlation coefficient of bonus contracts was not significantly different from that of the penalty contracts.

3 Conclusion

The study of neural substrates that are associated with economic risk has largely focused on lotteries. Yet, risk often appears in the economy in forms that are not comparable in substance to a lottery. Contractual obligations, which shift risk from principal to agent, are not strictly random events since the agent has considerable control of the outcome. Similar to lottery frames, contractual framing affected our participants' perceptions of the risk. In our study, participants behaviorally differentiated between penalty and bonus contractual frames. Participants purchased significantly more precautionary measures, i.e. the amount of additional questions purchased, in a contract framed as a penalty compared to those framed as a bonus. These findings agree with previous behavioral studies of contract framing (Hannan et al. 2005; Brink 2008).

Our neuroimaging results identify three regions which had significantly different BOLD signals in the contrast between bonus and penalty contracts during the active decision phase of the experiment. The motor-cortex showed significant differences, but given that participants used a button box to input their decisions, the difference in neural activity was likely the result of greater button usage. The motor cortex clusters were not correlated with the additional questions purchased but the clusters were correlated with the total number of button presses. The posterior insula clusters were also significantly different between the penalty and bonus framed contracts, but these regions were not correlated with the amount of additional questions purchased.

The dorsal striatum also showed significant differences between contractual frames. Moreover, the dorsal striatum correlated with the average additional questions purchased in each frame. The striatum has long been known to encode the value of outcomes, even when the outcome is uncertain. In particular, fMRI studies of risk and reward often implicated the ventral striatum in the valuation of outcomes (Knutson et al. 2001; Abler et al. 2006). The absence of a significant difference in the ventral striatum between the contract frames in either the passive contract phase or the active decision phase suggest that participants did not differentiate the values of the contracts. Previous studies of risk and reward revealed the outcome immediately after the lottery was presented, even if the participant was not paid until the end of the experiment or was randomly assigned a particular trial for payment. In our study, feedback for each contract was not presented while images were being acquired. Only after exiting the scanner and performing the task was feedback concerning their payoff presented. Therefore, participants should not have anticipated a reward during image acquisition, which may explain the absence of ventral striatum activity.

The interpretation that our participants were encoding both the cost of the action, i.e. the cost of an additional question, and the value of the contract, the cost of failure, would explain the negative correlation between activations in the dorsal striatum and the additional questions purchased. That is, if the participants are encoding the relative cost of an additional question with respect to the cost of failure, then a participant who perceived a higher cost of failure would have a lower valuation of the relative cost of an

additional question. Thus, participants who perceived a higher cost of failure would purchase more questions than participants who perceived a lower cost of failure.

Alternatively, the dorsal striatum could be encoding the probability of answering a trivia question correctly. Previous studies have shown that when probability and magnitude of a lottery are displayed separately, the probability was correlated with activity in the dorsal striatum while magnitude was correlated with activity in the ventral striatum (Berns and Bell 2012). If the dorsal striatum is encoding the probability of answering a trivia question correctly, then indeed the amount of additional questions purchased would be negatively correlated with the activations in the dorsal striatum. But the encoding of probabilities fails to explain why the framing of a contract's payoffs would elicit a difference in the perception of probabilities.

The lack of significant difference in neural activity between the bonus and penalty in the passive contract phase of the experiment suggests that our participants did not evaluate the contract without the context of the additional questions. It was only when the participants were making tradeoffs between the cost of additional questions and the risk of failing to complete the contract did we observe differences in the BOLD signal. The dorsal striatum seemed to play a significant role in differentiating between the frames since it showed not only a difference in activity between frames, but also the brain activity in the dorsal striatum is correlated with the amount of precaution, i.e. additional questions, purchased by the participant. This was contrary to previous studies of risk and framing effects where the ventral striatum or amygdala (Yacubian et al. 2006; Tom et al. 2007) displayed differences between positive and negative frames. The discrepancy between our study and studies using lotteries suggests that the evaluation of risk may be more task-dependent than previously thought. The extent to which the participants have control over the outcomes may be an important indication of which regions are recruited to evaluate a choice, but further experiments are required to determine if this is indeed the case.

Although we observed differences in both behavior and BOLD response to bonus and penalty contracts, it remains unclear whether loss aversion or overconfidence is responsible for the behavioral differences. We found no significant correlations between the loss aversion coefficients from our risk surveys and the difference in the amount of precaution purchased in the contractual frames, even when adjusting for nonlinear changes in probability for each additional question purchased. Additionally, none of the regions shown to correlate with loss aversion were active during our contracting task. Alternatively, the activity in the dorsal striatum correlated with the amount of precaution purchased and exhibited significantly more activity during the penalty contract compared to the bonus contract. As such, overconfidence is more likely responsible for behavioral differences in contractual frames.

The risk experienced by an agent in a principal-agent relationship seems to be fundamentally different from the risk experienced during a lottery. In particular, positive and negative framing experienced by an agent elicits different activity in the brain than the positive and negative framing experienced during a lottery. Although risk for both incentive contracts and lotteries evoke valuation in the striatum, the location within the striatum has been shown to correlate with different types of learning. The ventral striatum, which is linked to stimulus–response learning, is active in valuing a lottery while the dorsal striatum, which is linked to action-selection learning, is active in valuing an agent's effort. The difference in valuation could be related to the

difference in choice sets available to the contracted agent and an individual choosing between lotteries. For example, a gambler can only accept or reject a given lottery while agents can change the probability of success through precautionary measure, such as effort. But the perception of the risk is inseparable from the set of decisions available. Thus, standard models of risk for lotteries may not be applicable to risk experienced by agents.

References

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage, 31*(2), 790–795.

Allen, D. W., & Lueck, D. (1995). Risk preferences and the economics of contracts. *The American Economic Review, 85*(2), 447–451.

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *The Journal of Neuroscience, 27*(31), 8161–8165.

Berns, G. S., & Bell, E. (2012). Striatal topography of probability and magnitude information for decisions under uncertainty. *NeuroImage, 59*(4), 3166–3172.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience, 16*(13), 4207–4221.

Brink, A. (2008). The effects of risk preference and loss aversion on individual behavior under bonus, penalty, and combined contract frames. Electronic Theses, Treatises and Dissertations.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences, 1104*(1), 70–88.

Delgado, M., Locke, H., Stenger, V., & Fiez, J. (2003). Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 27–38.

Frederickson, J. R., & Waller, W. (2005). Carrot or stick? Contract frame and use of decision-influencing information in a principal-agent setting. *Journal of Accounting Research, 43*(5), 709–733.

Gervais, S., Heaton, J. B., & Odean, T. (2011). Overconfidence, compensation contracts, and capital budgeting. *The Journal of Finance, 66*(5), 1735–1777.

Goetz, C. J., & Scott, R. E. (1977). Liquidated damages, penalties and the just compensation principle: Some notes on an enforcement model and a theory of efficient breach. *C Law Rev, 77*, 554.

Hannan, R., Hoffman, V., & Moser, D. (2005). Bonus versus penalty: Does contract frame affect employee effort? In A. Rapoport & R. Zwick (Eds.), *Experimental business research* (Vol. II, pp. 151–169). New York: Springer.

Knutson, B., Fong, G., Adams, C., Varner, J., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport, 12*(17), 3683–3687.

Lazear, E. (1991). Labor economics and the psychology of organizations. *Journal of Economic Perspectives, 5*(2), 89–110.

Lewis, T. R. (1980). Bonus and penalties in incentive contracting. *Bell Journal of Economics, 11*(1), 292–301.

Luft, J. (1994). Bonus and penalty incentives contract choice by employees. *Journal of Accounting and Economics, 18*(2), 181–206.

Martino, B. D., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proceedings of the National Academy of Sciences, 107*(8), 3788–3792.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304*(5669), 452–454.

Penny, W., Flandin, G., & Trujillo-Barreto, N. (2007). Bayesian comparison of spatially regularised general linear models. *Human Brain Mapping, 28*(4), 275–293.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515–518.

Yacubian, J., Gläscher, J., Schroeder, K., Sommer, T., Braus, D. F., & Büchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *The Journal of Neuroscience, 26*(37), 9530–9537.