cbess
Centre for Behavioural and Experimental Social Science

University of East Anglia

2019-03

WHY DO WE LIE? DISTINGUISHING BETWEEN COMPETING LYING THEORIES

Paul Clist

Ying-yi Hong

www.uea.ac.uk/cbess

Why Do We Lie? Distinguishing Between Competing Lying Theories

Paul Clist[1] & Ying-yi Hong[2]

1. University of East Anglia
2. Chinese University of Hong Kong

July 2019

Lying is an important human behaviour that has received unprecedented empirical attention in recent years[1–4] due to a new experimental paradigm in which subjects are incentivised to misreport a die roll for financial gain.[5] Amongst theoretical attempts to explain results, Justified Dishonesty (JD) is a theory with impressive support and a plausible psychological foundation.[6–12] JD predicts subjects will swap a paid and unpaid roll of a die whenever financially beneficial, as this feels less like lying. However, JD's predictions are virtually identical to a competing economic model. In Dufwenberg & Dufwenberg's (DD)[13] subjects have perceived cheating aversion, incurring a cost of lying that is in proportion to the amount they are perceived to cheat. Current evidence is unable to distinguish between these two distinct mechanisms. Here we show that JD often makes accurate predictions, but is a poor explanation. We perform a placebo test, finding that JD is more accurate when it should be irrelevant. Furthermore, we elicit the second (unpaid) roll, strongly rejecting a direct corollary of JD. Our results demonstrate that the role of justifications and desired counterfactuals have been overstated. The simple idea that subjects dislike others perceiving them as liars in proportion to the size of the lie is sufficient to explain patterns of lying behaviour.

The reason humans lie can be trivial to explain: there is often advantage in doing so. Rather the puzzle is explaining why humans are often honest at a cost to themselves. Whilst initially promising, self-concept maintenance theory[14] has not survived replication attempts,[15] leaving two main camps. Current economic theories[1,2,16,17] typically rely on a combination of disliking lying and disliking being seen as a liar (see supplementary information, a2 for theoretical and practical differences between them). Justified Dishonesty is an alternative built on the psychological intuition that lies are less costly when there is an easy and ready justification.[6–12] In the supplementary info (a1, figure 3) we show that these distinct rationales can produce virtually identical predictions, if appropriately calibrated.

JD's mechanism is built on a common feature of dice rolling experiments: over 26,000 subjects have been asked to roll a die twice, before being asked to record their first roll.[1,5] This roll determines payment (with reports 1-5 receiving that number of monetary units, and 6 attracting no extra reward), while the second is neither paid nor reported. JD predicts subjects don't invent a roll they didn't experience, rather they merely report the higher of the two rolls. In observing counterfactuals, it is argued subjects are provided with ready justifications that feel less like lying.[6–11] This compelling idea results in intuitive predictions, with rolls of five to zero claimed in 11/36 to 1/36 cases, and an average mean claim prediction of 3.47.

The current evidence base for JD comprises three elements. First, the second roll increases mean claims,[6,7,11] with a meta analysis finding an effect size of between one fifth and one twentieth of a standard deviation.[1] Second, there is suggestive evidence on mechanisms,[6–9,11] with the switching of rolls being rated less negatively than outright lying. Third, JD's predictions of the mean and distribution of reports has been shown to be remarkably accurate. The most comprehensive test failed to reject JD in 13 of 23 countries,[12] showing wide applicability. However, the current evidence cannot disentangle the JD mechanism from competing theories. The impressive accuracy of JD's predictions is ambiguous, and could equally be interpreted as support for an appropriately calibrated economic model of lying (see figure 3).

We distinguish between the different types of lying theories using a two-fold strategy. First, we conduct a placebo test in Experiment 1 alongside a standard test in experiment 2. The placebo test involves eliciting subjects' first (paid) roll before asking subjects to roll a second time: JD should not be relevant as subjects have not observed any counterfactuals at the time of reporting their first roll. In experiment 2, we elicit the first roll after both rolls in the standard fashion.[5] JD should be relevant and accurate for experiment 2, but not for experiment 1.

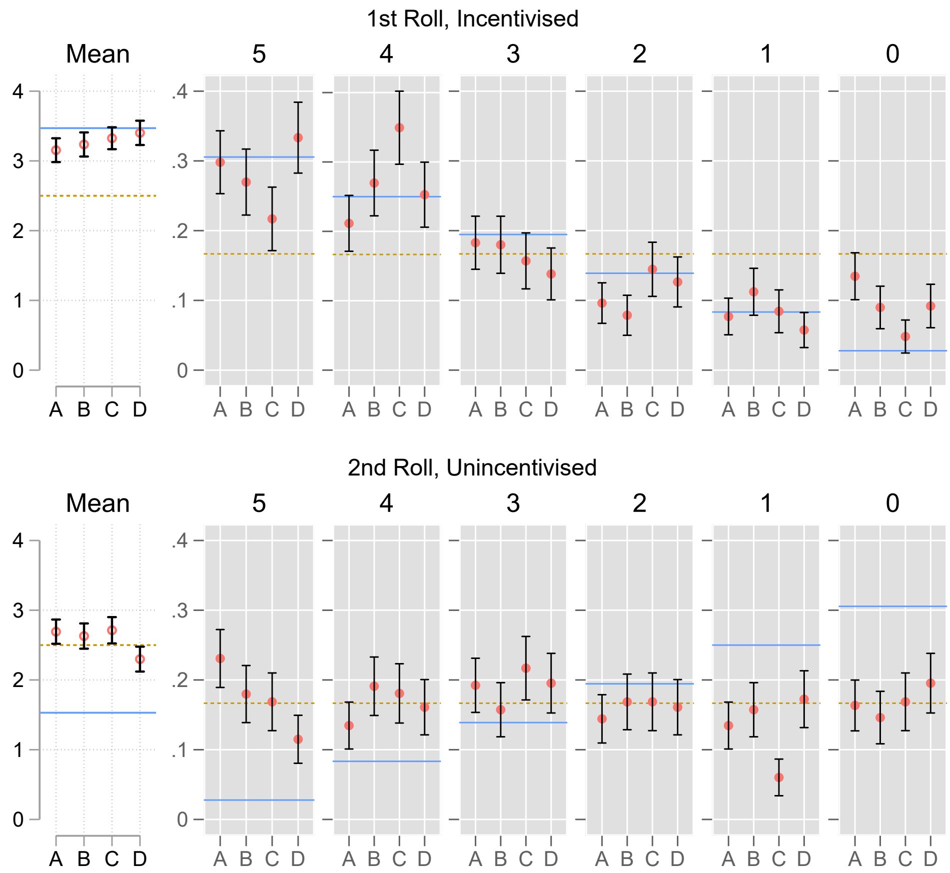
Second, we elicit the second (unpaid) roll, meaning we can test both halves of JD's story. A direct but untested corollary of JD is that if asked, subjects should report the lower roll as their second roll. In JD, the theory predicts no outright lies, rather subjects merely switch rolls. The predicted average report across the two rolls is the same as Full Honesty's prediction (2.5), but this comprises a higher first roll mean (3.47) and lower second roll mean (1.53). Regarding the distribution of second roll reports, these are likewise the mirror of the first roll. The distribution is reversed: subjects should report five in only 1/36 cases, with zero expected in 11/36 cases. By simply asking subjects to report both paid and unpaid rolls, we are able to test JD's mechanism.

In experiment 1, subjects ($n = 363$) in four groups across two countries roll and report two die rolls. The first roll determines payment, with the second roll encouraged to "...help us check that the dice we're using are fair". The main departure from the standard set up[5] is in recording the second roll, not in encouraging it. (See supplementary information for implementation details.) This is a placebo test because JD is not relevant for subjects that complete the survey in the order specified: they are only asked to roll a second time once they have already been asked to record their first. Behaviour should not conform to the JD's prediction in either roll.

The top half of figure 1 shows the mean claim and the distribution of claims for the first roll. The mean claim across the four groups is 3.27 ($sd = 1.6$), against predictions of 2.5 for Full Honesty and 3.47 for JD. We can easily reject Full Honesty as a prediction for all four groups, for both mean claim and the distribution ($p < 0.001$ in all eight cases). More surprisingly, given that it should not be relevant, JD is an accurate description of the mean and distribution of first round reports for all four groups. For one group, JD is rejected at the 10% level for both mean ($t = 1.85$, $p = 0.07$) and distribution (KSD's $d = 0.11$, $p = 0.054$). In the other 6 cases, the theory cannot be rejected even at the 10% level (the lowest p value is 0.17). For first roll behaviour, we cannot reject JD at the 5% level for any of the four groups for either mean or distribution of reports. This is stronger evidence in favour of JD than found by,[12] with JD proving accurate (when it shouldn't) for 4 distinct groups in two countries. The strength of evidence that JD is accurate for first roll behaviour implies subjects did perhaps follow JD by not completing the survey in the intended order. If JD is accurate for the paid roll, it should also be accurate for the unpaid roll.

For second roll behaviour, see the lower half of figure 1. The mean claim across the four groups is 2.59 ($sd = 1.7$), against predictions of 2.5 for Full Honesty and 1.53 for JD. Full Honesty is not rejected for any of the four groups using tests for mean or distribution (the lowest p value is 0.17). Full Honesty appears an accurate explanation of second roll behaviour, with actual claims within 0.05 standard deviations of the Full Honesty prediction. JD is strongly rejected for each of the four groups and for both mean and the distribution of reports ($p < 0.001$ in all eight cases). Descriptive statistics are also informative. While JD predicts no subject would report a lower roll for their first (incentivised) roll than for their second (unincentivised) roll, we see that 33% of subjects do so.

Figure 1: The Mean and Fraction at each report of both rolls, Experiment 1



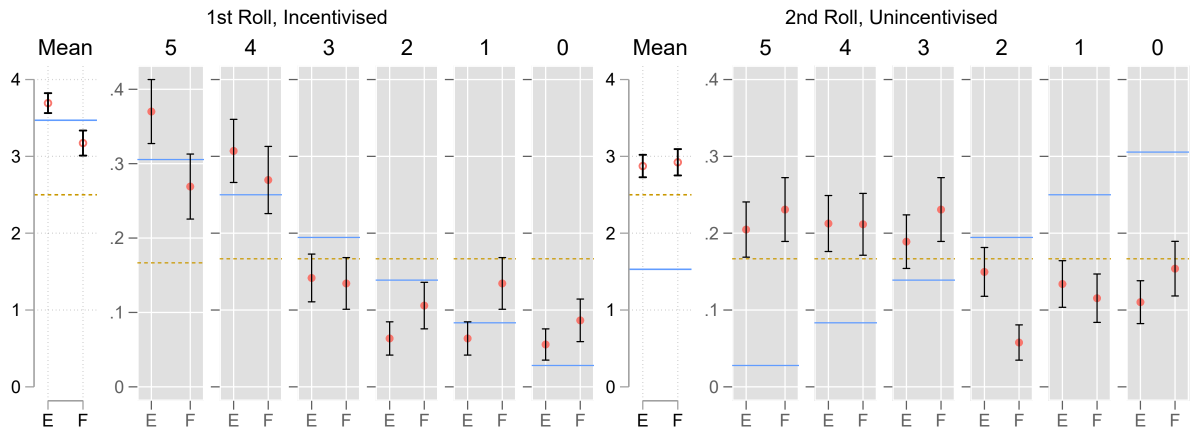
Note: Six, earning no extra reward, is coded as 0. Solid blue lines refer to the JD prediction. The dashed yellow line refers to the full honesty benchmark. Four samples are presented. A refers a sample of Chinese students at Beijing Normal University, China. B and C are both Chinese students at UEA, UK with B taking the survey in Chinese and C in English. D is a sample from UEA, UK (taking the survey in English). Sample differences in measured lying behaviour is much smaller than those reported in,[12] with no significant differences (one way ANOVA, F=0.42, p=0.74). Sample characteristics are irrelevant to this paper, but we keep the distinctions for clarity.

Experiment 1 is stronger evidence for the accuracy of JD predictions than:[6,12] we cannot reject JD for any of the four groups using the standard tests.[12] However, this was a placebo test: JD is an accurate prediction of lying behaviour when it should not be. If the theory performs most strongly in situations where it shouldn't be present, it begs the question as to whether justifications and counterfactuals are really causing the relevant behaviour in cases where JD should be present. The tests of second roll behaviour confirm that the underlying mechanism is not at work: evidence that the DD model[13] is a better explanation of first roll behaviour.

In experiment 2, we test JD in a setting in which it should be accurate. Subjects ($n =$ 231) report their first (paid) and second (unpaid) rolls in one of two treatments. In the 'two screen' treatment subjects are asked for their two rolls sequentially, on different screens of the computer-based experiment. In the 'one screen' treatment, subjects are asked simultaneously on a single screen. (Other protocols follow[12] closely, see supplementary information for more details.) The two treatments are found to be significantly different from each other in terms of the mean claim of the first roll (q=0.049), and so we present four tests of each hypothesis (mean and distribution, for each treatment). Our preanalysis plan[18] detailed all of the tests for the second experiment (and fixed those to be used on the first). For any theory (Full Honesty or JD) and roll, we report q-values using Hochberg's method, adjusting for the four tests conducted.

For first roll behaviour, the left panel of figure 2 shows the average report and distribution by treatment. The mean claim across the two treatments is 3.46: remarkably close to JD's prediction of 3.47. Despite this, tests reject JD at the 5% level due to the distribution of reports in the two screen treatment ($q = 0.04$, with $q = 0.09$ in the other three tests). Figure 2 shows the discrepancy, with JD underestimating the fraction of the sample reporting 5 or 4. We don't wish to over interpret this rejection of JD. The average mean claim is within 0.01 standard deviations, and only one of the four tests narrowly passes the 5% significance level we adopted in our pre-analysis plan. To give context, Full Honesty is easily rejected as a characterisation of first roll behaviour across the both treatments and tests (q<0.001 in all cases).

Figure 2: The Mean and Fraction at each report of both rolls, Experiment 2



Note: Solid blue lines refer to the JD prediction. The dashed yellow line refers to the full honesty benchmark. E and F refer to different treatments, both conducted at the UEA, UK. E refers to a two screen treatment, in which subjects reported their first roll on one screen, and were then asked to report their second on a different screen. F refers to a one screen treatment, in which subjects entered both numbers on a single screen.

For second roll behaviour, the right panel of figure 2 shows average reports and distribution. We can reject Full Honesty in all four cases (with q values ranging from 0.003 to 0.036). Much more strongly, all four tests reject JD as a description of second round reports ($q < 0.000000001$ in all cases). For second round behaviour, Full Honesty is statistically rejected, but is certainly the closer of the two theories. Descriptive statistics are again informative: 32% of subjects report a higher number for their unpaid roll than for their paid roll, while JD predicts no one would do this.

Our results reject JD as an explanation of lying behaviour: justifications and counterfactuals are not good explanations of lying behaviour. While JD is a reasonably accurate description of first roll behaviour, it is more accurate when it should be irrelevant. When JD should be present, it is strongly rejected for second roll behaviour ($p < 0.001$ in eight cases for experiment 1, and $q < 0.000000001$ in four cases for experiment 2). Amongst the competing economics models,[1,13,16,19] the percentage of partial liars shows DD[13] best fits the data (supplementary information, a2). DD is a psychological game theoretic model, which relies on the simple assumption that subjects dislike others perceiving them as liars in proportion to the size of the lie, which can produce virtually identical predictions to JD (supplementary information, a1). This can produce accurate predictions, and better explains lying behaviour.

References

[1] Abeler, J., Nosenzo, D. & Raymond, C. Preferences for truth-telling. *Econometrica* (2019).

[2] Jacobsen, C., Fosgaard, T. R. & Pascual-Ezama, D. Why do we lie? a practical guide to the dishonesty literature. *Journal of Economic Surveys* **32**, 357–387 (2018).

[3] Gerlach, P., Teodorescu, K. & Hertwig, R. The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin* **145**, 1–44 (2019).

[4] Rosenbaum, S. M., Billinger, S. & Stieglitz, N. Let's be honest: A review of experimental evidence of honesty and truth-telling. *Journal of Economic Psychology* **45**, 181–196 (2014).

[5] Fischbacher, U. & Föllmi-Heusi, F. Lies in disguise - an experimental study on cheating. *Journal of the European Economic Association* **11**, 525–547 (2013).

[6] Shalvi, S., Dana, J., Handgraaf, M. J. & De Dreu, C. K. Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes* **115**, 181–190 (2011).

[7] Shalvi, S., Eldar, O. & Bereby-Meyer, Y. Honesty requires time (and lack of justifications). *Psychological science* **23**, 1264–1270 (2012).

[8] Shalvi, S., Gino, F., Barkan, R. & Ayal, S. Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science* **24**, 125–130 (2015).

[9] Shalvi, S. Behavioural economics: Corruption corrupts. *Nature* **531**, 456 (2016).

[10] Bassarak, C. *et al.* What provides justification for cheating—producing or observing counterfactuals? *Journal of Behavioral Decision Making* **30**, 964–975 (2017).

[11] Gino, F. & Ariely, D. The dark side of creativity: original thinkers can be more dishonest. *Journal of personality and social psychology* **102**, 445 (2012).

[12] Gächter, S. & Schulz, J. F. Intrinsic honesty and the prevalence of rule violations across societies. *Nature* **531**, 496 (2016).

[13] Dufwenberg, M. & Dufwenberg, M. A. Lies in disguise–a theoretical analysis of cheating. *Journal of Economic Theory* **175**, 248–264 (2018).

[14] Mazar, N., Amir, O. & Ariely, D. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research* **45**, 633–644 (2008).

[15] Verschuere, B. *et al.* Registered replication report on mazar, amir, and ariely (2008). *Advances in Methods and Practices in Psychological Science* **1**, 299–317 (2018).

[16] Gneezy, U., Kajackaite, A. & Sobel, J. Lying aversion and the size of the lie. *American Economic Review* **108**, 419–53 (2018).

[17] Abeler, J., Becker, A. & Falk, A. Representative evidence on lying costs. *Journal of Public Economics* **113**, 96–104 (2014).

[18] Clist, P. & Hong, Y.-Y. The distribution of 1st and 2nd rolls in lying experiments: Distinguishing between competing theories. *AEA RCT Registry* https://www.socialscienceregistry.org/trials/3547/history/37865 (2018).

[19] Khalmetski, K. & Sliwka, D. Disguising lies-image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics* (2019).

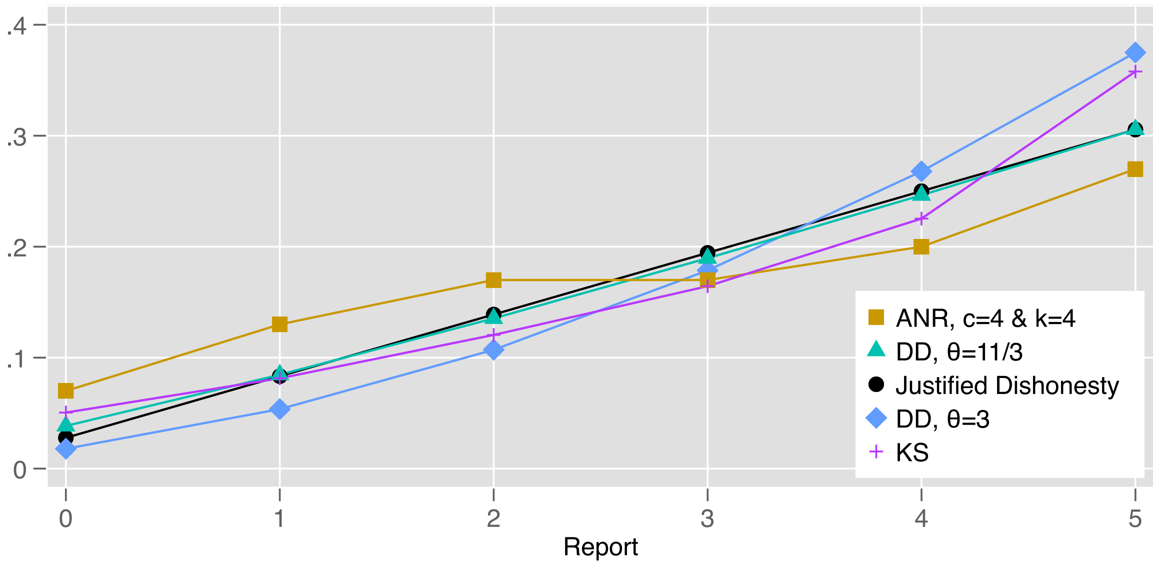
[20] Fischbacher, U. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics* **10**, 171–178 (2007).

A Supplementary Information

A.1 Calibration of the DD Model

The original DD model[13] is calibrated with $\theta = 3$, which gives qualitatively similar predictions to JD. However, the DD model is able to produce predictions that are virtually identical to JD. Using equation 11, we solve the model so that the proportion of subjects reporting 5 is the same as the JD model ($\Sigma_{x\leq 5}\pi_x \cdot s(x)(5) = 11/36$), which gives $\theta = 11/3$. The resulting distribution is shown in figure 3, with the largest discrepancy one percentage point. The average mean claim for JD is 3.47, against 3.44 for the newly calibrated DD model. Figure 3 also includes DD,[13] ANR[1] and KS[19] models as originally calibrated.

Figure 3: Five Predicted Distributions, from Four Models



A.2 Distinguishing DD from other economic models using partial lying

As shown in the main text, we are able to distinguish between JD and DD using placebo tests and recording the (unpaid) second roll. However, there are related lying models within economics, that also make similar predictions using different mechanisms. The calibrated predictions from two models[1,19] are shown in figure 3 ([19] is more tractable, and a special case of[16]). All three of these models combine a lying cost and a cost related to the chance of being seen as a liar. Differences between these models relate to how this second element is described, solution concepts and the understanding of heterogeneity of preferences/actions. One large dividing line with DD is that these three models assume additive separability: a lie is a fixed cost and not related to the lie's size.

Table 1: Partial and Maximal Lies, in Theory and Practice

	Theory				Experiment	
	JD[6]	DD[13]	ANR[1]	KS[19]	1	2
% lying	42	40	13	25	21.5*	28.1*
% maximal liars	14	14	10	19	11.4	15.8
% partial liars	28	26	3	6	10.0*	12.3*
Mean Claim	3.47	3.44	3.23	3.51	3.27	3.46

Note: $\star$, both the % of liars and the % of partial liars in the data is a minimum bound, rather than a point estimate. The percentage of maximal liars is simply the difference between the (full honesty) expected and actual number of subjects reporting the income-maximising report. The % of partial liars is then sum of all positive differences between all non-income maximising reports and the expectation of chance. This is a lower bound as the same distribution could be made up of a larger number of smaller lies. DD here is calibrated to match the JD prediction (i.e. $\theta = 11/3$). Experiment 1 and 2 are each presented as pooled data across their respective treatments.

This technical detail translates into large differences regarding the predictions of partial and full lying, as shown in table 1. This table shows that very different compositions of full and partial lies can give qualitatively similar mean claims. The existence and prevalence of partial lies in all three models are driven by the cost of being seen as a liar:[1,19] removing this element of the model means subjects either lie maximally or not at all. Our experimental data (shown in table 1) allows a judgement of whether lying costs are related to the size of the lie. While this cannot provide a clear estimate of the number of partial liars, it does produce a lower bound: both models that make clear predictions while assuming additive separability are well below this bound. This set of economic models is able to rationalise non-maximal lying, but they cannot rationalise the observed level of non-maximal lying. This is because the mechanism rests on the perceived chance of being a liar, the second derivative would need to be increasing over reports 3-5 in order for partial lying to be reported at each stage. The descriptive evidence presented above allows us to distinguish DD from other economics models which assume the cost of a lie is not directly related to its size. This evidence shows that social image or reputation effects (as currently modelled) are not sufficiently strong mechanisms to explain the prevalence of partial lying.

A.3 Experiment 1: Implementation Details

363 subjects participated in experiment 1 (67% female, mean age = 21.9, sd of age=3.2). The two rolls recorded here comprise the first section of a standalone survey experiment that took approximately 20 minutes to complete, conducted using paper and pen in classrooms where subjects were seated at least one (two-person) table apart with consistent

payments. This differs from,[5] which tends to be implemented after various other experiments, in lab settings, and on a computer[20] (though with real dice). For the British-based samples, we used exactly double the pay offfs of,[12] with a £2 show up fee and payments for dice rolls varying between £0 and £5. For Beijing we used a 10RMB=1GBP exchange rate, as this was both simple to administer and reflected PPP prices on campuses: the show up fee easily bought a hot drink, with the maximum additional pay-out enough to purchase a hot meal.

The English language survey for Chinese students is included below.

Figure 4: Survey

Please note: Talking is not allowed- please turn off your phone and do not talk to other students. If you cause a distraction, you will be asked to leave and will not earn any money. If you have any questions please raise your hand and a helper will come over as soon as possible.

Thank you for attending the session today, we really appreciate your time. Your participation is completely voluntary so you can leave at any time. Do be aware that if you do leave, you will forfeit any money you could have earned. For any specific questions which you don't wish to answer, just leave the answer blank.

This survey is part of joint work between Dr Paul Clist (paul.clist@uea.ac.uk) and Professor Hong Ying-Yi (yyhong@bnu.edu.cn). If you have any concerns you are welcome to contact them by email. We're interested in how students from China make decisions, and so we'll ask you a few brief questions. Many of the questions are taken from the World Values Survey or the Global Preference Survey. For these questions we are interested in what you prefer and there is no 'correct' choice.

Your answers here will be completely confidential. You may start whenever you are ready to do so.

1. As you are aware, you will be paid for your help today with the exact amount due to chance. Please roll the dice in front of you, recording the amount below.

	☐	☐	☐	☐	☐	☐
Number on the dice:	1	2	3	4	5	6
This is worth:	£1	£2	£3	£4	£5	£0
Show up fee:	£2	£2	£2	£2	£2	£2
Total:	£3	£4	£5	£6	£7	£2

2. To help us check that the dice we're using are fair, roll the dice once more (this won't affect your pay off) and record the dice roll here ____________________

A.4 Experiment 2: Implementation Details

231 subjects participated in experiment 2 (56% female, average age = 21.4, sd of age=3.3). All sessions were run at UEA's experimental lab in Norwich, UK using the standard experimental software.[20] Protocols followed the original[5] set up, with the experiment being run directly after another experiment, and pay outs of 0-£2.50 matching.[12]

In both treatments, the first screen of the experiment is identical to.[5] The visual presentation is identical throughout, so here we only reproduce the text, with any departures in bold. In each case, the number denotes the screen number.

1. For the following questionnaire you will receive a small additional payoff. However, this payoff is not the same for every participant. You determine your own payoff by throwing your die twice as soon as you are asked to.
   Your first throw decides on how much you receive. You can see the exact payoff from the following chart. It will remain on the screen until you have entered your throw.
   The second throw only serves to make sure that the die is working properly. You may of course throw the die more than twice. However, only the first throw counts. If you have any question, please raise your hand. If you are ready, please press OK

2. Please throw the die **twice** now.
   Please keep in mind the first **and second numbers** you have thrown.
   If you have thrown the die **twice**, please press OK.

Here, the two treatments diverge.

Treatment E:

3. Now please enter the **first number you have thrown**.
   1st number thrown: (with box to enter)
   Resulting payoff: (with box that updates)

4. *Without payoff table*
   Now please enter the second number you have thrown.
   2nd number thrown: (with box to enter)

Treatment F:

3. *Two boxes, side by side. Box 1:*
   Now please enter the first number you have thrown.
   1st number thrown: (with box to enter)
   Resulting payoff: (with box that updates)

   Box 2, on the right
   Now please enter the second number you have thrown.
   2nd number thrown: (with box to enter)

A.5 All Results

Here we include the results of all tests that were described in the pre-analysis plan.[18]

Table 2: Experiment 1: Mean and Distribution Tests of 1st & 2nd Roll Behaviour Against two leading benchmarks

		Summary		Mean		Distribution	
Theory's Prediction	Sample	Mean	sd	t	p	KSD	p
JDB: $\bar{r}_1 = 3.47$	a	3.15	1.74	1.85*	0.07	0.11*	0.05
	b	3.24	1.63	1.35	0.18	0.09	0.17
	c	3.34	1.45	0.92	0.36	0.09	0.20
	d	3.40	1.62	0.39	0.70	0.06	0.42
FHB: $\bar{r}_1 = 2.5$	a	3.15	1.74	3.83***	0.0002	0.19***	0.0001
	b	3.24	1.63	4.26***	0.0001	0.22***	0.0000
	c	3.34	1.45	5.25***	0.0000	0.23***	0.0000
	d	3.40	1.62	5.18***	0.0000	0.25***	0.0000
JDB: $\bar{r}_2 = 1.53$	a	2.69	1.77	6.68***	0.0000	0.31***	0.0000
	b	2.63	1.71	6.07***	0.0000	0.28***	0.0000
	c	2.71	1.69	6.28***	0.0000	0.34***	0.0000
	d	2.30	1.66	4.31***	0.0000	0.22***	0.0001
FHB: $\bar{r}_2 = 2.5$	a	2.69	1.77	1.11	0.27	0.06	0.43
	b	2.63	1.71	0.71	0.48	0.04	0.89
	c	2.71	1.69	1.13	0.26	0.10	0.17
	d	2.30	1.66	1.13	0.26	0.06	0.58

Note: Sample sizes are 104 (A), 89(B), 83(C) and 87(D). 3 subjects in C did not report a second roll.

Table 3: Tests of Treatment Equality

	Summary Statistics						Tests of					
	e (2 screens)			f (1 screen)			Mean			Distribution		
	N	$\bar{r}$	r_{sd}	N	$\bar{r}$	r_{sd}	T	p	q	W2	p	q
Roll 1	127	3.69	1.46	104	3.17	1.66	2.53**	0.01	0.05	7.12	0.13	0.39
Roll 2	127	2.87	1.65	104	2.92	1.75	0.22	0.83	0.83	3.21	0.52	0.83

Note: As specified in the preanalysis plan, the four tests are used in the hochberg procedure to produce q values. ***, ** and * denote $q \leq 0.01$, $q \leq 0.05$ and $q \leq 0.1$ respectively.

Table 4: Experiment 2: Mean and Distribution Tests of 1st & 2nd Roll Behaviour Against two leading benchmarks

Theory's		Summary		Mean			Distribution		
Prediction	Treatment	Mean	sd	t	p	q	KSD	p	q
JDB: $\bar{r}_1 = 3.47$	e	3.69	1.46	1.72*	0.09	0.09	0.12**	0.01	0.04
	f	3.17	1.66	1.82*	0.07	0.09	0.11*	0.04	0.09
FHB: $\bar{r}_1 = 2.5$	e	3.69	1.46	8.97***	0.00	0.00	0.34***	0.00	0.00
	f	3.17	1.66	3.94***	0.00	0.00	0.21***	0.00	0.00
JDB: $\bar{r}_2 = 1.53$	e	2.87	1.65	9.19***	0.00	0.00	0.36***	0.00	0.00
	f	2.92	1.75	8.12***	0.00	0.00	0.42***	0.00	0.00
FHB: $\bar{r}_2 = 2.5$	e	2.87	1.65	2.56**	0.01	0.03	0.11**	0.04	0.04
	f	2.92	1.75	2.47**	0.02	0.03	0.17***	0.00	0.00

Note: There are four hypothesis (two theories making predictions about two rolls). Each is tested four times, using each treatment and both the distribution and the mean. Hochberg's q values are reported, adjusting for the four tests of each prediction.

A.6 Funding

The preparation of this paper was partially supported by a General Research Fund (Ref. No. 14655416) from the Research Grant Council of Hong Kong SAR government and a research grant from Beijing Normal University awarded to Ying-Yi Hong, and grants from DEVCo (UEA) and CBESS awarded to Paul Clist.

A.7 Authors Contributions

Paul Clist co-wrote the survey and experiment, oversaw the UK data collection, analysed the data, calibrated the DD model and wrote the article. Ying-Yi Hong co-wrote the survey and experiment, translated the survey into Chinese and oversaw the Beijing data collection.

A.8 Competing Interests

The authors declare no competing interests.