ONLINE APPENDICES

Appendix A: Robustness Checks

The empirical specification used in the main text of this paper assumes that updating follows the flexible parametric process described in Equation 1. This specification allows for a wide range of deviations from Bayes' rule, as discussed in Section 2. In this section we conduct several exercises to test for the robustness of the main results.

The first subsection examines whether the results from the main specification described in Equation 2 are robust to first differencing the dependent variable (i.e. this considers how new information influences the change in beliefs, imposing the assumption that $\delta = 1$). The second subsection extends the main empirical specification to allow for individual-specific updating parameters. The third subsection pools all the observations across the three treatments together, and then tests whether the average updating parameters differ across treatments, by interacting treatment group dummies with the regressors of the main specification described in Equation 2.

Table 4: First Difference Specification and Power Calculations

	T1: SYMMETRIC			T2: COMBINED			T3: SEPARATE		
	OLS	IV	DIFF	OLS	IV	DIFF	OLS	IV	DIFF
	(1a)	(1b)	(1c)	(2a)	(2b)	(2c)	(3a)	(3b)	(3c)
δ	0.90	0.99		0.86	0.99		0.93	0.99	
	(0.03)***	(0.03)		(0.04)***	(0.02)		(0.02)***	(0.02)	
γ_a	1.09	1.09	1.09	1.06	1.02	1.01	1.16	1.14	1.13
	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)
$\gamma_b - \gamma_a$	0.08	0.08	0.08	0.07	0.10	0.11	-0.03	-0.03	-0.02
	(0.08)	(0.08)	(0.08)	(0.10)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
p($\gamma_a = \gamma_b$)	0.32	0.31	0.32	0.48	0.25	0.25	0.73	0.74	0.75
MDE ($\kappa = 0.8$)	0.24	0.22	0.22	0.27	0.25	0.26	0.24	0.22	0.22
R^2	0.73		0.31	0.74		0.21	0.84		0.34
1st Stage F		84.04			107.01			95.45	
N	1,075	1,075	1,075	1,285	1,285	1,285	1,140	1,140	1,140

(i) Standard errors in parentheses (clustered at the individual level)

(ii) T-tests of H_0: $\delta = 1$; $\gamma_a = 1$; $\gamma_b - \gamma_a = 0$ indicated by * = 10%, ** = 5%, *** = 1%

(iii) MDE reports the minimum detectable effect size for a power of κ.

Robustness Check 1: First-Differences Specification and Power Calculation

This section of the robustness checks serves two purposes. The first purpose is to check for the robustness of the results from the core empirical specification to the use of a first differences specification (DIFF), which essentially involves imposing the assumption that $\delta = 1$. The second purpose of this section is to report the size of the minimum detectable effect (MDE) from power calculations for both our main OLS and IV empirical specification, and the DIFF specification.

One of the challenges in carrying out a statistical analysis of belief updating behavior is that an individual's current posterior belief necessarily depends upon her prior belief, which in turn is the result of updating in response to past information. Therefore, when estimating a parametric belief updating function, one concern is that the individual's prior belief is correlated with unobservables. In the main text, we devoted substantial space to discussing how the experiment was designed explicitly to address this concern by generating a completely exogenous information set, facilitating a natural instrumental variables (IV) approach to estimation. The first differences specification results presented here serve to further complement the IV analysis, since the DIFF specification avoids the potential endogeneity issue by removing the lagged belief from the set of dependent variables in the regression.

In columns (#a) and (#b), Table 4 repeats the OLS and IV results from Table 3 for the corrected beliefs, with one minor change to the core specification in Equation 2. Here we report, instead, the results for the equivalent specification:

$$\tilde{\pi}_{i,j,t+1} = \delta\tilde{\pi}_{i,j,t} + \gamma_a\widehat{q} - (\gamma_b - \gamma_a)\tilde{q} \cdot 1(s_{i,j,t+1} = b) + \epsilon_{i,j,t+1} \tag{3}$$

where $\tilde{\pi}_{i,j,t} = \text{logit}(\pi_{i,j,t})$ and $\widehat{q} = \log(\frac{q}{1-q}) \cdot [1(s_{i,j,t} = a) - 1(s_{i,j,t} = b)]$; while as above, $\tilde{q} = \log(\frac{q}{1-q})$; j refers to a round of decisions; t counts the decision numbers within a round, and the errors ϵ_{ijt+1} are clustered at the individual (i) level. The difference $\gamma_b - \gamma_a$ denotes a single parameter estimated in the regression, but is denoted as the difference between γ_b and γ_a as this is the natural way to think about this parameter in the context of the discussion above (i.e. the difference between how subjects update in response to 'bad news' and 'good news').

The reason for the rearrangement of the equation is that, while it is equivalent[21] to the specification in Equation 2, it displays the test of the difference between γ_a and γ_b more clearly (i.e. the

[21]Notice that the regression coefficients and standard errors on δ and γ_a are the same in Tables 3 and 4 (where we are only considering the corrected beliefs). Furthermore, we can see the equivalence from the following simple rearrangement:

$$\begin{aligned}
\tilde{\pi}_{i,j,t+1} &= \delta\tilde{\pi}_{i,j,t} + \gamma_a\tilde{q} \cdot 1(s_{i,j,t+1} = a) - \gamma_b\tilde{q} \cdot 1(s_{i,j,t+1} = b) \\
&= \delta\tilde{\pi}_{i,j,t} + \gamma_a\tilde{q} \cdot 1(s_{i,j,t+1} = a) - \gamma_a\tilde{q} \cdot 1(s_{i,j,t+1} = b) + \gamma_a\tilde{q} \cdot 1(s_{i,j,t+1} = b) - \gamma_b\tilde{q} \cdot 1(s_{i,j,t+1} = b) \\
&= \delta\tilde{\pi}_{i,j,t} + \gamma_a\tilde{q} \cdot [1(s_{i,j,t+1} = a) - \tilde{q} \cdot 1(s_{i,j,t+1} = b)] + [\gamma_a - \gamma_b] \cdot \tilde{q} \cdot 1(s_{i,j,t+1} = b)
\end{aligned}$$

test of the *asymmetric updating hypothesis*), and thereby also facilitates calculating the MDE. In Table 3, we have presented the MDE for a power of $\kappa = 0.8$.

Columns (#c) report the results for the first difference specification, which imposes the restriction that $\delta = 1$:

$$\Delta\tilde{\pi}_{i,j,t+1} = \gamma_a \widehat{q} - (\gamma_b - \gamma_a)\tilde{q} \cdot 1(s_{i,j,t+1} = b) + \epsilon_{i,j,t+1} \tag{4}$$

where $\Delta\tilde{\pi}_{i,j,t+1} = \text{logit}(\pi_{i,j,t+1}) - \text{logit}(\pi_{i,j,t})$ and $\widehat{q} = \log(\frac{q}{1-q}) \cdot [1(s_{i,j,t} = a) - 1(s_{i,j,t} = b)]$; j refers to a round of decisions; t counts the decision numbers within a round, and the errors ϵ_{ijt+1} are clustered at the individual (i) level.

The results indicate that the $\gamma_b - \gamma_a$ parameter is robust to the different empirical specifications adopted, and also doesn't vary substantially across treatment groups. In all treatment groups, and for each of the empirical specifications considered, we cannot reject the null hypothesis that this parameter is equal to zero, which implies that we do not find support for the *asymmetric updating hypothesis*. Furthermore, we calculate the MDE for each specification, considering a significance level of $\alpha = 0.05$ and a power of $\kappa = 0.8$. Under these assumptions, the MDE for the difference between the γ_b and γ_a parameters in each of the regressions considered in isolation ranges between 0.22 and 0.27. As a result, we cannot conclusively reject the possibility that there exists a small asymmetry in updating; however none of our results provide any support for this conclusion.

Robustness Check 2: Allowing for Individual-Specific Updating Parameters

As discussed above, one reason we might think that endogeneity of the lagged belief could lead to biased estimates is if there is heterogeneity in individual updating behavior and this leads to a correlation between the unobserved error term and the lagged belief variable amongst the regressors. We have tried to address this issue above using, firstly, an instrumental variable approach, and secondly, a first differences empirical specification. However, since the data were collected in the form of a panel of belief updates for each individual, the data lends itself to controlling for individual-specific behavior through exploiting the panel. A typical fixed effects model is not appropriate here, as it is not the *level* of the regression that shifts from individual to individual. However, we can include individual-specific updating parameters to control for the *slope* to shift at the individual level. This allows us to extract the individual heterogeneity in how responsive individuals are to their prior belief, and to new information in general, and reduce the possible bias in the main parameter of interest, the average difference in responsiveness to 'bad news' and 'good news' : $\gamma_b - \gamma_a$. With this in mind, our third robustness check involves estimating the following empirical specification:

$$\tilde{\pi}_{i,j,t+1} = \delta_i \tilde{\pi}_{i,j,t} + \gamma_i \hat{q} - (\gamma_b - \gamma_a)\tilde{q} \cdot 1(s_{i,j,t+1} = b) + \epsilon_{i,j,t+1} \tag{5}$$

where δ_i and γ_i are estimated at the individual level, and the remaining parameters and variables are defined as above. The results from this exercise using the corrected beliefs are reported in Table 5.

These results are very consistent with the estimates from the core specification, as well as from the DIFF specification in Robustness Check 2. In summary, all the empirical estimates provide support for the same underlying story that the data collected in this experiment provide no support for the *asymmetric updating hypothesis* in this context.

Table 5: Allowing for Individual-Specific Updating Parameters.

	T1 SYMMETRIC	T2 COMBINED	T3 SEPARATE	T2+T3
	(1)	(2)	(3)	(4)
$\gamma_b - \gamma_a$	0.09	0.14	-0.04	0.06
	(0.10)	(0.10)	(0.09)	(0.07)
p($\gamma_a = \gamma_b$)	0.35	0.18	0.67	0.41
MDE ($\kappa = 0.8$)	0.28	0.29	0.26	0.20
MDE ($\kappa = 0.9$)	0.33	0.33	0.30	0.23
N	1,075	1,285	1,140	2,425
R^2	0.80	0.84	0.89	0.86

(i) Standard errors in parentheses

(ii) T-tests of H_0: Coefficient = 0 reported: * = 10%, ** = 5%, *** = 1%

(iii) MDE reports the minimum detectable effect size for a power of κ.

Robustness Check 3: Between Treatments Comparison of Updating Parameters

This section tests whether the belief updating parameters in our core specification are significantly different *between* the three treatment groups. This is done by pooling together the three treatment groups and estimating Equation 2, but with the inclusion of treatment dummies interacted with the updating coefficients. This provides us with a test of whether the parameters differ between either of the two ASYMMETRIC treatments and SYMMETRIC.

More specifically, this involves estimating the following equation:

$$\tilde{\pi}_{i,j,t+1} = \delta\tilde{\pi}_{i,j,t} + \gamma_a\tilde{q}\cdot 1(s_{i,j,t+1} = a) - \gamma_b\tilde{q}\cdot 1(s_{i,j,t+1} = b) + \sum_{k=2}^{3}[\delta^k\tilde{\pi}_{i,j,t}\cdot T^k_{i,j,t} + \gamma^k_a\tilde{q}\cdot 1(s_{i,j,t+1} = a)\cdot T^k_{i,j,t} - \gamma^k_b\tilde{q}\cdot 1(s_{i,j,t+1} = b)\cdot T^k_{i,j,t}] + \epsilon_{i,j,t+1}$$

where $T^k_{i,j,t}$ is an indicator variable for treatment k [i.e. $T^k_{i,j,t} = 1(T_{i,j,t} = k)$], with $T_{i,j,t}$ a treatment variable taking the values $\{1, 2, 3\}$ corresponding to the three treatment groups. The coefficients δ,γ_a, and γ_b reflect the baseline parameters without the influence of state-contingent stakes and the parametersδ^k, γ^k_a and γ^k_b estimate the movement from these parameters for each of the two state-contingent stake treatments, $k \in \{2, 3\}$.

The results from this exercise are presented in Table 6. The results show that, for the average individual, there are no systematic differences in the updating parameters across treatment groups. This implies that the differences in exogenous state-contingent incentives do not exert a strong influence on how individuals update their beliefs in the different treatments.

Table 6: Testing for Differences in Average Updating Behavior *between* Treatment Groups.

	Belief 1 (1)	Belief 2 (2)	Belief 3 (3)	Belief 4 (4)	Belief 5 (5)	Pooled (6)	Full Sample (7)
Priors							
δ	0.93	0.89	1.03	1.00	1.02	0.99	1.00
	(0.05)	(0.04)	(0.05)	(0.06)	(0.06)	(0.03)	(0.03)
δ * T2	-0.02	0.02	-0.02	0.01	0.01	-0.00	-0.04
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.04)	(0.03)
δ * T3	0.02	0.06	-0.10	0.08	-0.04	0.00	-0.00
	(0.08)	(0.06)	(0.06)	(0.07)	(0.09)	(0.04)	(0.03)
Signal: Blue ($s = a$)							
γ_a	0.81	1.02	1.03	1.30	1.28	1.09	0.84
	(0.14)	(0.18)	(0.18)	(0.21)	(0.21)	(0.11)	(0.09)
γ_a * T2	0.09	-0.07	-0.15	-0.08	-0.16	-0.07	-0.01
	(0.22)	(0.26)	(0.22)	(0.34)	(0.29)	(0.15)	(0.13)
γ_a * T3	0.12	-0.05	-0.10	0.07	0.21	0.05	0.04
	(0.21)	(0.22)	(0.20)	(0.28)	(0.31)	(0.15)	(0.13)
Signal: Red ($s = b$)							
γ_b	0.69	0.79	1.18	1.40	1.80	1.16	0.78
	(0.11)	(0.16)	(0.19)	(0.21)	(0.23)	(0.11)	(0.09)
γ_b * T2	0.21	0.06	-0.33	-0.21	0.14	-0.04	0.11
	(0.20)	(0.22)	(0.25)	(0.27)	(0.40)	(0.17)	(0.15)
γ_b * T3	0.30	0.04	-0.15	-0.37	-0.06	-0.05	0.10
	(0.18)	(0.20)	(0.21)	(0.25)	(0.33)	(0.15)	(0.12)
Observations	700	700	700	700	700	3,500	5,550
Kleibergen-Paap F	46.61	47.46	56.63	60.45	64.03	93.17	53.88

(i) Robust standard errors in parentheses (clustered at the individual level).
(ii) Estimates use the corrected beliefs and are instrumented using the correct lagged Bayesian posterior.
(iii) All of the non-interacted coefficients are significantly different from 0 at the 1% level. Only one of the forty-two interaction coefficients are significantly different from zero at the 10% level. This is the γ_b * T3 coefficient in the Belief 1 column, which is significant at the 10%, but not the 5% level.

Appendix B.1: Core Properties of Bayes' Rule

Möbius et al. (2014) argue that the core structure of Bayesian updating is captured by the following three properties:

1. *invariance,* whereby the difference in logit beliefs between t and $t+1$ depends only on the history of signals, H_{t+1}, and the initial prior, p_0 (i.e. on the agent's information set). An updating process is *invariant* if we can find a function g_t such that:

$$\text{logit}(\pi_{t+1}) - \text{logit}(\pi_t) = g_t(s_{t+1}, s_t, ..., s_1; p_0)$$

   If an individual displayed base rate neglect or confirmatory bias, this would constitute a violation of invariance (i.e. in the context of the model outlined in the main text, this assumption stipulates that $\delta = 1$).

2. π_t is a *sufficient statistic* for all information received at time t or earlier, such that the change in logit beliefs depends only on the new information in time $t+1$: $\text{logit}(\pi_{t+1}) - \text{logit}(\pi_t) = g_t(s_{t+1})$

3. *stability* of the updating process over time. This property is satisfied if $g_t = g$ for all t.

Under the assumption that these properties are satisfied, the authors note that the class of updating processes that remain can be fully described by the two parameter function, $g(s_t)$, where:

$$g(s_t) = \log(\frac{q}{1-q}) \cdot 1(s_{t+1} = a) - \log(\frac{q}{1-q}) \cdot 1(s_{t+1} = b)$$

This serves to motivate the model described in Equation 1.

Appendix B.2.1: The QSR and a Non-EU 'Truth Serum'

In this section, we discuss how beliefs reported under the QSR might be distorted, and how we address this challenge. Consider the binary event, denoted by E_ω, where $\omega \in \{A, B\}$. Therefore, E_A refers to the event that state $\omega = A$ is realized. The object that we would like to elicit is the participant's belief, $\pi_t = P(E_A) = P(\omega = A)$, regarding the likelihood that state $\omega = A$ is the correct state at time t. However, the object that we will observe is the participant's reported belief, r_t, at each point in time under the incentives prescribed by the quadratic scoring rule. The *Quadratic Scoring Rule* at time t is defined by:

$$S_A(r_t) = 1 - (1 - r_t)^2 \quad (6)$$

$$S_B(r_t) = 1 - r_t^2 \quad (7)$$

where r_t is the reported probability of event E_A occurring; $S_A(r_t)$ is the payment if the state $\omega = A$ is realized; $S_B(r_t)$ is the payment if the state $\omega = B$ is realized. Therefore, the QSR essentially involves a single choice from a list of binary prospects, $(1-(1-r_t)^2)_{E_A}(1-r_t^2)$. The QSR is a '*proper*' scoring rule since, if the agent is a risk neutral EU maximizer then she is incentivized to truthfully reveal her belief, π_t:

$$\pi_t = arg \max_{r_t \epsilon [0,1]} \pi_t S_A(r_t) + (1-\pi_t) S_B(r_t)$$

However, the QSR is no longer incentive compatible once we allow for (i) *risk aversion / loving* and (ii) participants who have exogenous stakes in the state of the world. The reasons for this are the following. Firstly, it has been well documented theoretically that, if the participant is *risk averse,* then the QSR leads to reporting of beliefs, r_t, that are distorted towards 0.5, away from her true belief, π_t, when the participant has no exogenous stakes in the realized state.[22] This distortion has been observed in experimental data (Offerman et al., 2009; Armantier and Treich, 2013). Secondly, in our experiment, we are also interested in eliciting beliefs when participants have an exogenous stake associated with one of the two states. More precisely, we are interested in recovering the participant's true belief when she receives an exogenous payment, x, if state $\omega = A$ is realized. This payment, x, is in addition to the payment she receives from the QSR. In other words, she chooses from a menu of binary prospects of the form: $(x+1-(1-r_t)^2)_{E_A}(1-r_t^2)$.

In the context of state-dependant stakes, a risk averse EU maximizer[23] faces two distortionary motives in reporting her belief: (i) she faces the motive to distort her belief towards 0.5 as discussed above; and (ii) in addition, there is a hedging motive, which will compel a risk averse individual to *lower* her reported belief, r_t, towards zero as x increases.

If the participants in our experiment are *risk neutral expected utility* maximizers, the reported beliefs, r_t, that we elicit under the QSR will coincide with their true beliefs, π_t. However, in order to allow for choice behaviors consistent with a wider range of decision models, we measure the size of the distortionary influence of the elicitation incentives at an individual level and correct the

[22] i.e. if $\pi_t > 0.5$ then $\pi_t > r_t > 0.5$, and if $\pi_t < 0.5$ then $\pi_t < r_t < 0.5$ for a risk averse individual reporting her beliefs under QSR incentives.

[23] A participant who is a risk averse EU maximizer chooses her reported belief r_t by solving the following maximization problem:

$$\max_{r_t \epsilon [0,1]} \pi_t U(x + 1 - (1-r_t)^2) + (1-\pi_t) U(1 - r_t^2)$$

beliefs accordingly. This approach is valid under the weak assumption that individuals evaluate binary prospects according to the *biseparable preferences*[24] model and are *probabilistically sophisticated*.[25] This restriction on behavior is very weak and includes individuals who behave according to EU with any risk preferences as well as the majority of commonly used NEU models.[26]

A Non-EU 'Truth Serum'

The discussion above has highlighted how beliefs might be distorted under QSR incentives. The Offerman et al. (2009) approach proposes correcting the reported beliefs for the risk aversion caused by the curvature of the utility function or by non-linear probability weighting. This approach involves eliciting participants' reported belief parameter, r, for a set of risky events where they know the objective probability, p (*known probability)*. This is done under precisely the same QSR incentive environment in which we elicit the participants' subjective beliefs, π, regarding the events of interest (where they don't know the objective probability: *unknown probability*). If a subject's reported beliefs, r, differ from the known objective probabilities, p, this indicates that the subject is distorting her beliefs due to the incentive environment (e.g. due to risk aversion). The objective of the correction mechanism is therefore to construct a map, R, from the objective beliefs, $p \in [0, 1]$, to the reported beliefs, r, for each individual under the relevant incentive environment.

Offerman et al. (2009) show that under the assumption that individuals evaluate prospects in a way that is consistent with the weak assumptions of the *biseparable preferences* model, then in the scenario where there are no state-contingent stakes (i.e. $x = 0$), individuals evaluate the QSR menu of prospects $(1-(1-r_t)^2)_{E_A}(1-r_t^2)$ according to $w(P(E_A))U(1-(1-r_t)^2) + (1 - w(P(E_A)))U(1-r_t^2)$ for $r_t \geq 0.5$ and therefore the inverse of the map from objective probabilities

[24]The *biseparable preference* model holds if the preference ordering, $\succsim$, over prospects of the form, $y_E z$, can be represented by:

$$y_E z \to W(E)U(y) + (1 - W(E))U(z)$$

where U is a real-valued function unique up to level and unit; and W is a unique weighting function, satisfying $W(\varnothing) = 0$, $W(S) = 1$ and $W(E) \leq W(F)$ if $E \subseteq F$. S is the set of all states and events are subsets of the full set of states: i.e. $E, F \subseteq S$. In this paper, we only consider two-state prospects, where the state-space is partitioned into two parts by an event, E and its complement E^c. Making the further assumption that the decision maker is *probabilistically sophisticated* gives the following refinement:

$$y_E z \to w(P(E))U(y) + (1 - w(P(E)))U(z)$$

[25]*Probabilistic sophistication* is the assumption that we can model that individual's preferences over prospects as if the individual's beliefs over states can be summarized by a probability measure, P. In other words, probabilistic sophistication implies that we can model the individual's belief regarding the likelihood of an event E as being completely summarized by a single probability judgment, $P(E_A)$.

[26]Amongst the models subsumed within the biseparable preferences model are EU, Choquet expected utility (Schmeidler, 1989), maxmin expected utility (Gilboa and Schmeidler, 1989), prospect theory (Tversky and Kahneman, 1992), and α-maxmin expected utility (Ghirardato et al., 2004). See Offerman et al. (2009) for a discussion.

to reported probabilities, R, is given by:

$$p = R^{-1}(r) = w^{-1}\left(\frac{r}{r + (1-r)\frac{U'(1-(1-r)^2)}{U'(1-r^2)}}\right) \tag{8}$$

In the next section, we provide a derivation for this equation, as well as augmenting the Offerman et al. (2009) approach to allow for the scenario where there are state-contingent stakes (i.e. $x \neq 0$). This extension to Offerman et al. (2009) represents a special case of the more general treatment of correction methods for binary proper scoring rules considered by Kothiyal et al. (2011). In our empirical analysis, we discuss how we use Equation 8 to recover the function, R, for each individual and thereby recover their beliefs, π_t, from their reported beliefs, r_t.

Appendix B.2.2: Augmenting the Offerman et al. (2009) 'Truth Serum' Approach to Include Stakes

The previous section discussed the central ideas motivating the Offerman et al. (2009) approach for correcting for hedging in cases where there are no state-dependent stakes (i.e. $x = 0$). In projects studying the *asymmetric updating hypothesis*, allowing for state-dependent stakes (i.e. $x \neq 0$) is of fundamental importance. Therefore, in this section, we consider an extension to the Offerman et al. (2009) approach to correcting for hedging. The extension we consider is tailored specifically to our experimental setting, however it is a special case of the more general set of correction techniques studied by Kothiyal et al. (2011).[27]

In the case where $x \neq 0$, the text above discussed how participants who face the quadratic scoring rule incentives, along with the non-zero state-contingent bonus x, essentially face a choice from a menu of lotteries denoted by $(x + 1 - (1-r_t)^2)_{E_A}(1 - r_t^2)$. An individual who satisfies the *biseparable preferences* model and is probabilistically sophisticated will evaluate this prospect using the following Equations:[28]

For $x \geq 1$ or $r_t \geq 0.5$:

$$w(P(E_A))U(x + 1 - (1-r_t)^2) + (1 - w(P(E_A)))U(1 - r_t^2) \tag{9}$$

[27]Kothiyal et al. (2011) extend the basic idea used by techniques aiming to correct elicited beliefs for reporting bias (e.g. hedging) to apply to the set of all binary proper scoring rules, and cover the full domain of beliefs. In conjunction with Offerman et al. (2009), this paper therefore offers a useful set of tools for accessing subjects' true beliefs in situations where they may have reason to distort their reports.

[28]For expositional simplicity, we don't consider $x \in (0, 1)$. The discussion below is easily extended to these cases, but they are irrelevant for the purposes of this paper. This case is slightly different due to the fact that the probability weights on events or states may depend on their ordinal ranking according to preferences in this model.

and similarly,

For $x = 0$ & $r_t < 0.5$:

$$(1 - w(P(E_A^c)))U(x + 1 - (1 - r)_t^2) + w(P(E_A^c))U(1 - r_t^2) \tag{10}$$

The reason for the two separate conditions is due to the way in which many NEU models, subsumed within *biseparable preferences* model, allow the probability weighting function, $w(.)$, over events to be influenced by the ordinal ranking over the associated outcomes, from best to worst.[29] Since the case where $x = 0$ is discussed extensively in Offerman et al. (2009), we will focus on the case where $x \geq 1$ in the discussion that follows. This case only requires a very minor adjustment to their discussion. The key results are the following (adjusted to include the influence of x):

Result 1: *Under NEU with* ***known probabilities****, p, the optimal reported probability, $r = R_x(p)$ satisfies:*

$$\text{If } x \geq 1\text{,then } p = R_x^{-1}(r_t) = w^{-1}\left(\frac{r_t}{r_t + (1 - r_t)\frac{U'(x+1-(1-r_t)^2)}{U'(1-r_t^2)}}\right) \tag{11}$$

Result 2: *Under NEU with* ***unknown probabilities****, the optimal reported probability, r, satisfies:*

$$\text{If } x \geq 1\text{,then } P(E) = w^{-1}\left(\frac{r_t}{r_t + (1 - r_t)\frac{U'(x+1-(1-r)_t^2)}{U'(1-r_t^2)}}\right) \tag{12}$$

This motivates the simple strategy for recovering the agent's subjective beliefs from her reported beliefs under the specific incentive environment that she faces. Since the RHS of (11) and (12) agree, we have:

$$P(E) = R_x^{-1}(r) \tag{13}$$

which implies that if we can recover the function R_x^{-1} then we can map the reported beliefs to the participant's subjective beliefs. Equation 11 shows that we can recover this R_x^{-1} function in the same way here, with the bonus payment of x, as in the case where $x = 0$ considered in Offerman et al. (2009). Essentially, we provide the participant with prospects over known probabilities, p, and ask them for their belief regarding the likelihood that one state will be realized. In order to ensure the incentives to distort one's reported beliefs are kept constant, we do this exercise under precisely the same incentive environment as in the main belief updating task. By eliciting these

[29] When $x = 0$ and $r_t < 0.5$, then $1 - r_t^2 > 1 - (1 - r_t)^2$ (i.e. in this case, E_A^c becomes the preferred event, rather than E_A, and therefore the probability weighting function is reversed).

reported beliefs associated with known probabilities spanning the whole unit interval, we can use these (p, r_t) pairs to estimate $R_x(p)$ for each individual for the relevant incentive environment created by the belief elicitation. Having estimated $R_x(p)$, we can calculate its inverse, $R_x^{-1}(r)$.

We can take any beliefs reported by the participant under the same belief elicitation incentives and then use this estimated $R_x^{-1}(r)$ to recover her true beliefs. In particular, we can use this estimated function to recover her true beliefs from her reported beliefs in the belief updating task that is the focus of this paper. Essentially, we are using this procedure to remove any misreporting effect that the belief elicitation incentive environment may have. It allows us to correct for the possibility that individuals may hold some belief $P(E)$ or π, but instead report a different belief, r.

If the incentive environment does not cause the participant to report a belief different from her true belief, then this procedure is unnecessary, but applying the procedure to her reported beliefs will not have any effect. In this case, the corrected beliefs will be the same as the reported beliefs.

Appendix B.2.3: Calibration of the Belief Correction Procedure: Theory

It is clear from the discussion in the main text and Equation 8, that we could recover $R(.)$ *non-parametrically* for each individual if we were to collect a large number of (p, r) pairs from participants, such that the interval between the known probabilities, p, is sufficiently small. However, since it is not practical here to elicit such a large number of observations from each participant, we instead impose a parametric structure similar to the one used by Offerman et al. (2009).

For the utility function, $U(.)$, we use the *constant relative risk aversion (CRRA)* functional form:

$$U(x) = \begin{cases} x^\rho & \text{if } \rho > 0 \\ \ln x & \text{if } \rho = 0 \\ -x^\rho & \text{if } \rho < 0 \end{cases} \tag{14}$$

For the probability weighting function, $w(.)$, we adopt Prelec's (1998), one-parameter family:[30]

[30]This is a special case of Prelec's two-parameter family of weighting functions:

$$w(p) = \exp[-\beta(-\ln(p))^\alpha]$$

For the purposes of the current context, the two-parameter family is not practically suitable due to the limited data we use at the individual level. This functional form permits the standard inverse-S shaped probability weighting function that has been found to be consistent with the majority of the existing empirical evidence. When $\beta = 1$ in the one-parameter family, the α parameter captures the degree of curvature of the inverse-S shape but the point at which

$$w(p) = \exp[-(-\ln(p))^{\alpha}] \tag{15}$$

Substituting these parametric functional form specifications into Equation 8 for the case where $x = 0$ gives:

$$p = R^{-1}(r_t) = \exp\left(-\left[-\ln\left(\frac{r_t(2r_t - r_t^2)^{1-\rho}}{r_t(2r_t - r_t^2)^{1-\rho} + (1-r_t)(1-r_t^2)^{1-\rho}}\right)\right]^{\frac{1}{\alpha}}\right) \tag{16}$$

For the the case where $x \neq 0$, substituting these parametric functional form specifications described in Equations 14 and 15 into Equation 11 gives:

$$p = R_x^{-1}(r_t) = \exp\left(-\left[-\ln\left(\frac{r_t(1-(1-r_t)^2 + x)^{1-\rho}}{r_t(1-(1-r_t)^2 + x)^{1-\rho} + (1-r_t)(1-r_t^2)^{1-\rho}}\right)\right]^{\frac{1}{\alpha}}\right) \tag{17}$$

We therefore use this adapted specification for our correction mechanism for the COMBINED and SEPARATE treatment groups.

In our core analysis, for our individual level reported belief corrections, we will make the simplifying assumption that $\alpha = 1$, such that risk aversion is captured only through the curvature of the utility function and not through the probability weighting function. The results are similar when we use Prelec's one parameter weighting function. Furthermore, it is substantially easier to interpret the risk aversion parameter estimates when we estimate ρ alone, due to the strong relationship between the ρ and α estimates .

We therefore estimate the following model, for each participant, in order to acquire a numerical estimate for the inverse of this function, $R(.)$:

$$\text{logit}[R(j/20)] = \text{logit}[h(j/20, \alpha, \rho)] + u_j \tag{18}$$

where $R(j/20)$ is the probability reported by the individual that corresponds to true known probability, $p = \frac{j}{20}$ where $1 \leq j \leq 19$.[31] As discussed above, α is the parameter of the probability weighting function; ρ gives the curvature of the utility function. The function $h(.)$ is the inverse of R^{-1}. We estimate this function, $h(.)$ numerically at each step[32] within the maximum likelihood estimation. The error terms, u_j, are independently and identically distributed across participants and choices and are drawn from a normal distribution. Essentially, here we are using each participant's 20 (r, p) pairs in order to estimate an R function that reflects the distortion in her reported

$w(p)$ intersects the 45 degree line is predetermined. Adding the second parameter, β, extends the one-parameter specification by allowing this fixed point to vary.

[31] In other words, for known probabilities, p, between 0.05 and 0.95 at intervals of 0.05.

[32] i.e. given the current parameter guesses.

beliefs due to the particular quadratic scoring rule incentive structure that she is subject to. Notice, that this structure varies across treatments as x varies and therefore the same subject would require a different adjustment curve if she were reassigned to a different treatment.

Using these estimates at an individual level allows us to recover the participants' true subjective beliefs, π_t, from the first stage of the experiment in which they report their beliefs, r_t, regarding the likelihood of $\omega = A$ being the true state. In Figure 6, we graph the individual level correction curve estimates for two individuals in each treatment group. It is clear from these examples, firstly, that individuals in the sample are distorting their reported beliefs substantially relative to the known probabilities, and secondly, that the estimated correction curves are sufficiently flexible to fit different types of belief distortion behavior reasonably well. Furthermore, importantly, the graph in the top-left panel of the figure shows that, when an individual accurately reports her beliefs, then the correction mechanism has no harmful effect.

At the aggregate level, for each treatment group, $T \in \{1, 2, 3\}$, we estimate:

$$\text{logit}[R_i(j/20)] = \text{logit}[h_i(j/20, \alpha, \rho)] + u_{i,j} \tag{19}$$

where j indexes the 20 reported probabilities of individual i. This specification allows us to examine the distortion caused by the incentive environment to the average individual in each of the three treatment groups.

Appendix B.2.4: Calibration of the Belief Correction Procedure: Estimation

The belief correction procedure that we adopt involves assuming a flexible parametric form for the participants' utility and probability weighting functions in order to estimate the R function discussed in Equation 8 above. We estimate this function for each individual separately in order to correct the reported beliefs at the individual level. In addition, we estimate this function at the aggregate level for each of the treatment groups in order to obtain a measure of the average distortion of the incentive environment faced in each of the treatment groups. A detailed discussion of the mechanics of the Belief Correction Procedure we use is provided in Appendix B.2.3 above. Essentially, we are simply fitting a curve through each subject's belief elicitation incentive distortion.

Figure 5 displays the average correction curves for each of the treatment groups, fitting a single curve to the reported belief data observed across all subjects in the relevant treatment group. Comparing the three subgraphs, we see that the average individual distorts the beliefs she reports in a way that is consistent with what risk aversion under EU would predict, with the inverse-S shape distortion in the T1.Symmetric stakes treatment and the strong distortion downwards (away from the more desirable state) in both of the Asymmetric stakes treatments. Furthermore, we see that the different ways of framing the same incentives in the two Asymmetric treatments has a clear influence on behavior. In T2.Combined, the participants hedge far more when choosing their reported beliefs in comparison to those in the T3.Separate group. This is in spite of the fact that the incentives are identical in these two treatments. This indicates that the reported beliefs in the T3.Separate treatment are closer to the participants' true beliefs and motivates this presentation of incentives as preferable for future work that calls for the elicitation of beliefs when there are exogenous state-contingent payments.

At the individual level, there is a large degree of heterogeneity in the degree to which individuals distorted their reported belief away from their actual belief, given the incentive environment. Figure 6 displays the correction curves estimated for two individuals from each treatment group. It is clear from this figure that some individuals responded very strongly to the incentive environment in which their belief was elicited, while others reported their belief more accurately. The belief correction procedure is therefore very helpful for recovering the true beliefs of participants who responded strongly to the incentive environment. In cases where the individual simply reported their belief accurately, the corrected beliefs and the reported beliefs are exactly the same.

Figure 5: Average Correction Functions across Treatments.

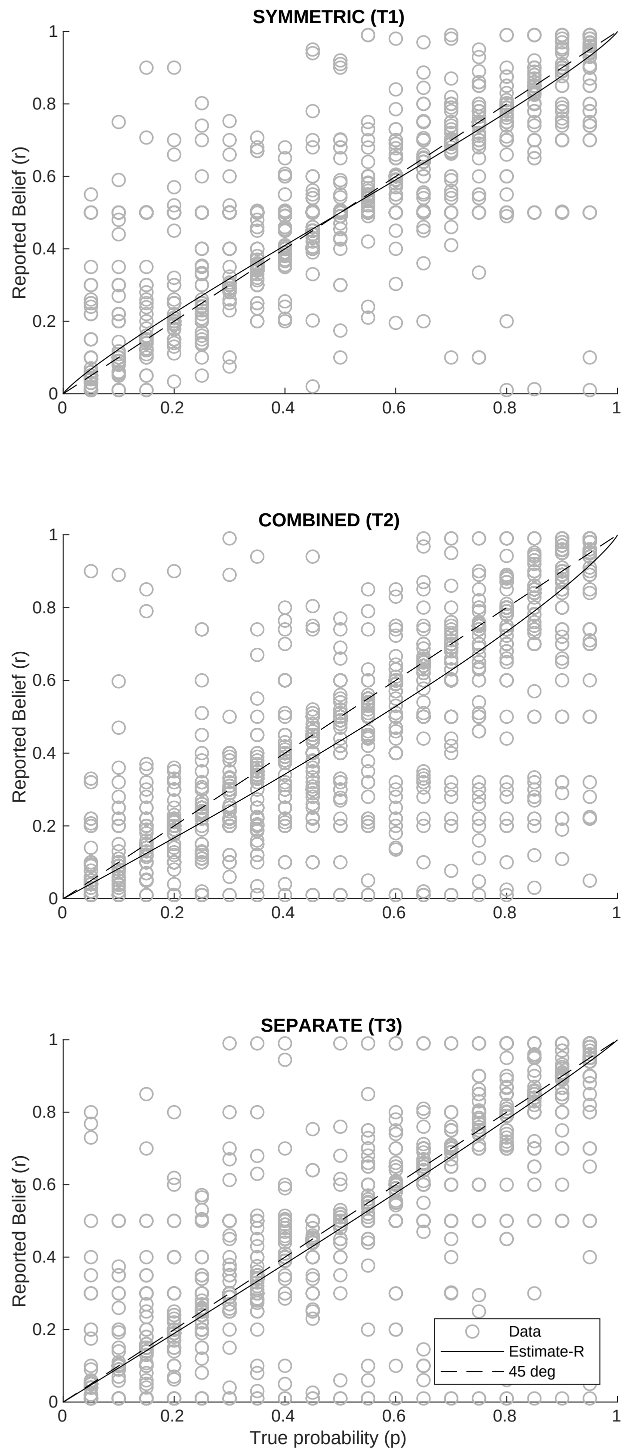
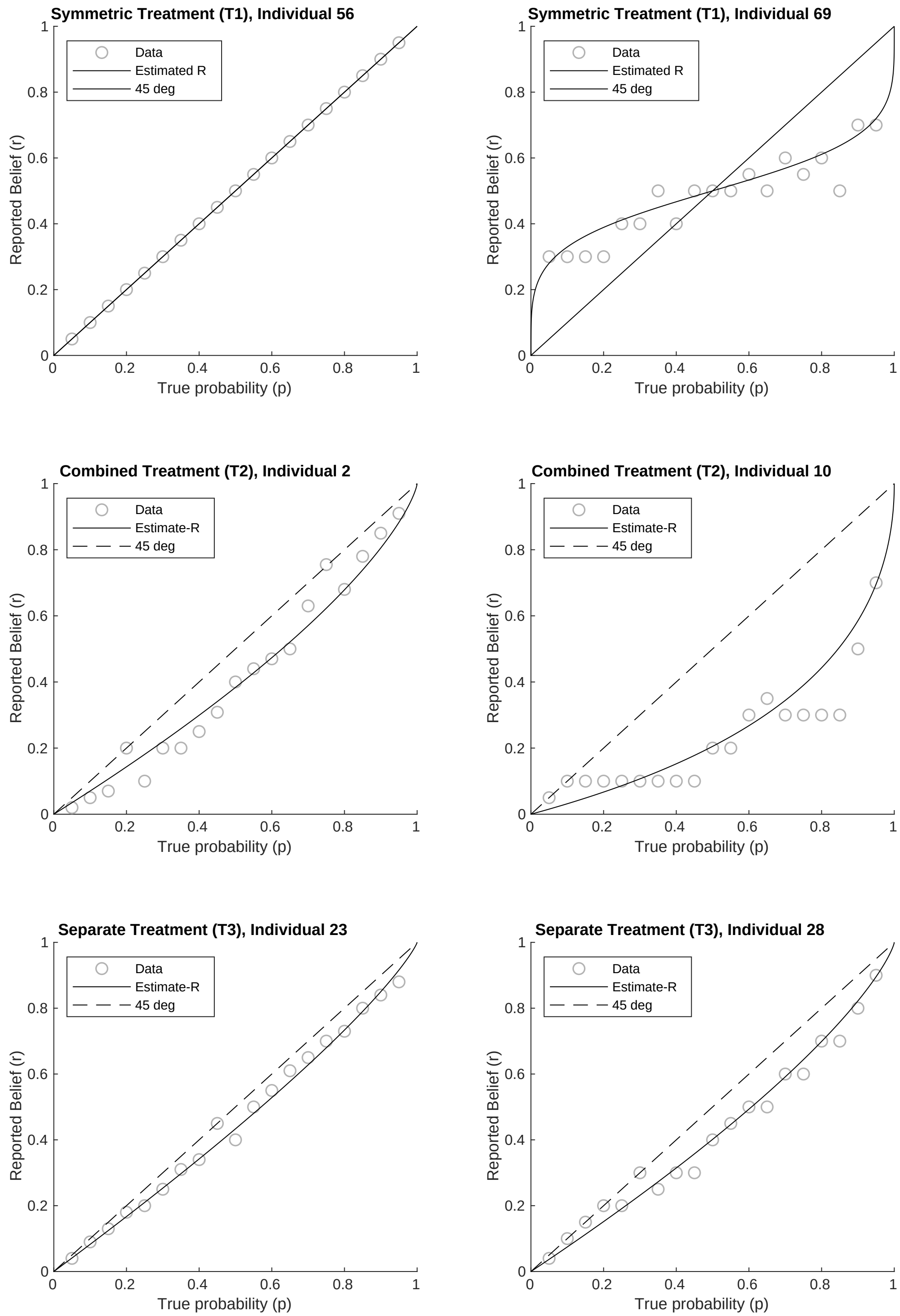


Figure 6: Individual Level Estimates of the Incentive Correction Function.



Appendix C: Endogeneity—An Illustrative Example

This section provides a simple illustrative example of why it is important to exercise caution in addressing possible endogeneity issues when studying belief updating. When considering belief updating in real world scenarios, we are often interested in studying how individuals update their beliefs from home-grown prior beliefs (i.e. subjective prior beliefs that are not exogenously endowed to subjects). However, when studying how subjects update beliefs from a home-grown prior, it is important to pay careful attention to the possible endogenous relationships between: (i) individual updating types, (ii) states of the world, and (iii) prior beliefs.[33] In the illustrative example I discuss below, I consider the implications of a relationship between (i) and (ii), allowing for (iii) to be completely exogenous.

One domain where studying belief updating from home-grown priors is essential is the domain of beliefs about the *self*. Beliefs about one's self deserve special attention since they are of critical importance in guiding our interaction with the world around us. This set of beliefs are also of central importance to the literature considering the *good-news, bad-news* hypothesis, since these beliefs are often heavy in affect, and are amongst the beliefs that we care most about.

However, when we study belief updating from subjective priors, there is a danger that an individuals prior beliefs are related to the way she updates her beliefs. Furthermore, if the beliefs pertain to her *self*, her prior belief may be related to this fundamental, implying that the distribution of signals she receives is related to this fundamental and to her prior. This is true even if the individual receives "exogenous" noisy signals about the fundamental.

The following discussion has the objective of illustrating one possible way in which neglecting to pay attention to the endogeneity of the signal distribution can be problematic. The example is purely hypothetical, and rather contrived, but serves to illustrate the basic point. In particular, I use a very simple "toy" simulation to demonstrate that ignoring this issue can (in principal) lead to mistakenly find evidence for asymmetric updating when all individuals update symmetrically. It is important to point out that I am *not* suggesting that this is the explanation for asymmetric updating results observed in the literature[34]—I am simply highlighting a potential endogeneity issue that should be addressed in this literature going forwards. In this regard, I also suggest a simple solution for dealing with the issue.

Much of the belief updating literature considers situations that resemble the following basic structure: consider an agent who updates about two states of the world, $\omega \in \{High, Low\}$, and receives a sequence of noisy signals, $s_t \in \{UP, DOWN\}$. This also reflects the setup considered in the

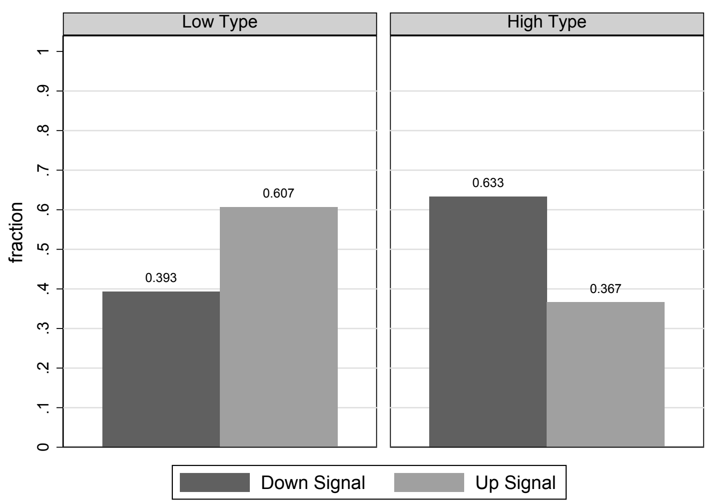
[33] In some cases, e.g. when forming beliefs about the self, there are differences across individuals in the fundamental (ii) that they are forming beliefs about. In these cases, the relationship between (i) and (iii) might be mediated by a natural relationship between the fundamental (ii) and the priors.

[34] For example, Figure 3 in Möbius et al. (2014) suggests that this is probably not a major concern for their main results. However, it is still important to control for this potential endogeneity issue as a robustness check.

current paper. However, since it is very important to also understand how we form beliefs about the *self,* in some studies the states are determined by personal characteristics of the individual (e.g. IQ). This means that states are essentially equivalent to personal types (i.e. *states* = *types*). The implication of this is that if signals are informative about the state of the world, then *High* types are more likely than *Low* types to receive *Up* signals (and vice versa for *Down* signals). If *High* types update their beliefs differently from *Low* types, this can (in principal) lead to finding (what looks like) evidence that the average individual updates asymmetrically when no individual actually does.

In order to show this, I conduct a very simple simulation exercise. I construct a population of 10 000 individuals who are randomly assigned to one of two types, $\omega \in \{High, Low\}$. Within each type, the agents' prior beliefs about the likelihood of being the *High* type are assigned randomly using a uniform distribution, distributed between zero and one.[35] *High* types receive an *Up* signal with probability $q = \frac{5}{8}$ and *Low* types receive a *Down* signal with probability $q = \frac{5}{8}$. Using a seed of 1000 in STATA, the observed empirical distribution of signals across types is shown in Figure 7.

Figure 7: Frequencies of Signals by Type



Now, the important part of this story is that belief updating may (in principal) be related to the underlying fundamental of interest. For example, it is conceivable that high IQ individuals process information and update their beliefs differently from low IQ individuals.[36]

[35]Note, this is an unrealistic assumption. In general, prior beliefs are related to the true state of the world. For example, beliefs about one's rank in an IQ distribution tend to be correlated with one's actual rank. However, for the purposes of this illustration, constructing type and prior belief to be orthogonal allows us to isolate only the effect of the endogeneity of types and signals (with exogenous priors). Allowing for priors to be related to the underlying fundamental would add an additional layer of endogeneity issues.

[36]Note, even if one doesn't find the story that the two *types* (here, *High* and *Low*) might update their beliefs differently compelling, a very similar pattern could also be generated if there is a relationship between *prior beliefs* and

Here, we consider two types that use different, *but always symmetric*, updating rules. In particular, we consider a *High* type that is perfectly Bayesian, and a *Low* type that is not very responsive to new information (but otherwise very well behaved her belief updating).[37]

The *High* type updates according to the following rule ($\delta = 1$, $\gamma_{UP} = 1$, $\gamma_{DOWN} = 1$):

$$\text{logit}(\pi_{t+1}) = 1 \cdot \text{logit}(\pi_t) + 1 \cdot \log(\frac{5}{3}) \cdot 1(s_{t+1} = UP) - 1 \cdot \log(\frac{5}{3}) \cdot 1(s_{t+1} = DOWN) \quad (20)$$

The *Low* type updates according to the following rule ($\delta = 1$, $\gamma_{UP} = 0.2$, $\gamma_{DOWN} = 0.2$):

$$\text{logit}(\pi_{t+1}) = 1 \cdot \text{logit}(\pi_t) + (0.2) \cdot \log(\frac{5}{3}) \cdot 1(s_{t+1} = UP) - (0.2) \cdot \log(\frac{5}{3}) \cdot 1(s_{t+1} = DOWN) \quad (21)$$

However, if we as the analyst neglect the possibility that the two types update their beliefs differently, then we might obtain biased parameters. This is illustrated by the regression estimates presented in column 1 and 2 of Table 7 below. These columns reflect the estimates from the standard specification used in this literature (i.e. equation 2). These parameter estimates, along with the true population averages are summarised as follows:

	True Parameter Values (Population Ave.)	Estimates
δ	1	1
γ_{UP}	0.6	0.7
γ_{DOWN}	0.6	0.5

It is clear from this that in spite of the fact that there is not a single individual in this population who updates asymmetrically that the estimated parameters suggest that there is an asymmetry. Notice, the standard errors are small and the adjusted R^2 suggests a good model fit (see Table

updating. This follows because: (i) prior beliefs about one's self are typically related to the underlying fundamental in question (e.g. beliefs about one's IQ positively correlated with actual IQ), (ii) types are mechanically related to the distribution of signals in the class of experiments we're considering, and therefore (iii) prior beliefs are related to the distribution of signals observed. Therefore, the story described in this section is worth paying attention to in any situation where at least one of the following might be violated:

$$(1) \text{ states of the world} \perp \text{belief updating and } (2) \text{ priors} \perp \text{belief updating}$$

[37]Note, neither of the types makes any errors in their belief updating. They both follow their updating rule perfectly. This exercise therefore rules out several other channels that can make life challenging for the analyst (e.g. errors related to priors or types).

7). As mentioned above, this is simply an illustration of why it is important to pay attention to the relationship between the distribution of signals and the types. Furthermore, it is important to point out that one can make an similar argument to show that even if the majority of individuals in the population are asymmetric updaters that the neglect of a relationship between signals and types could (in principal) generate estimates that suggest symmetric updating.

Fortunately, this particular endogeneity issue is easy to deal with by simply considering updating behavior within each type (e.g. interacting the RHS variables of equation 2 with the Type dummy variable). This is illustrated in column 3 of Table 7. Notice, also, that simply including the Type dummy variable in the regression does not solve the problem (see column 2).

Table 7: Estimates of Simulated Data Parameters

	Model 1 (1)	Model 2 (2)	Model 3 (3)
δ	1.002 (0.001)	1.002 (0.001)	1.000 (.)
γ_{UP}	0.696 (0.005)	0.683 (0.007)	0.200 (.)
γ_{DOWN}	0.504 (0.006)	0.513 (0.006)	0.200 (.)
High Type (=1)		0.0114 (0.004)	
High Type (=1) * δ			0.000 (.)
High Type (=1) * γ_{UP}			0.800 (.)
High Type (=1) * γ_{DOWN}			0.800 (.)
N	10000	10000	10000
Adjusted R^2	0.99	0.99	1.00

(i) Standard errors in parentheses
(ii) Note: Std errors in column 3 missing due to perfect fit.

Appendix D: Supplementary Figures, Results and Experimental Instructions

Figure 8: Comparison of Initial (Period 0) Belief with the Exogenous Prior

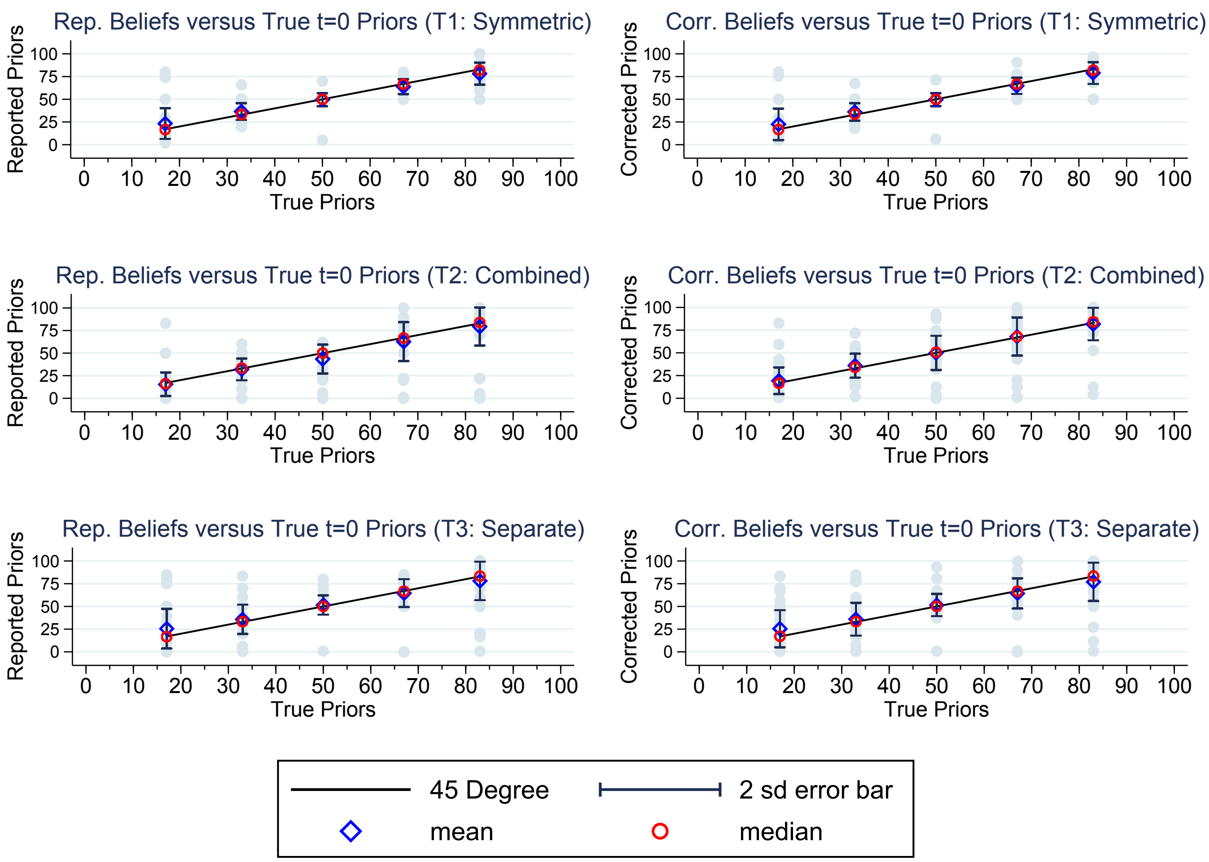


Table 8: First Stage Regressions Output Associated with Table 3

	T1: SYMMETRIC		T2: COMBINED		T3: SEPARATE	
	Reported	Corrected	Reported	Corrected	Reported	Corrected
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
δ_{IV}	0.81	0.82	0.95	0.92	0.88	0.85
	(0.06)***	(0.05)***	(0.06)***	(0.06)***	(0.06)***	(0.05)***
γ_a	-0.07	-0.07	-0.60	-0.05	-0.03	-0.09
	(0.11)	(0.10)	(0.22)***	(0.20)	(0.25)	(0.25)
γ_b	-0.17	-0.20	0.34	-0.20	-0.24	-0.21
	(0.10)*	(0.11)*	(0.20)*	(0.17)	(0.23)	(0.26)
Kleibergen Paap F	170.39	228.03	238.80	256.15	220.64	276.53
Shea Partial R^2	0.58	0.61	0.48	0.50	0.50	0.49
N	1,075	1,075	1,285	1,285	1,140	1,140

(i) Standard errors in parentheses (clustered at the individual level).

Table 9: Testing for Sample Balance Across Treatment Groups

	T1 SYMMETRIC	T2 COMBINED	T3 SEPARATE	D(2,1)	D(3,1)	D(3,2)
Gender (Male = 1)	0.56	0.58	0.58	0.02	0.02	0
	(0.50)	(0.50)	(0.50)			
Age	22.88	23.02	21.88	0.14	-0.99	-1.13
	(3.46)	(4.47)	(3.11)			
Location (London = 1)	0.51	0.48	0.52	-0.03	0.01	0.04
	(0.50)	(0.50)	(0.50)			
Economics Class	0.44	0.54	0.42	0.10	-0.02	-0.12
	(0.50)	(0.50)	(0.50)			
Home Language German	0.42	0.45	0.42	0.03	-0.00	-0.03
	(0.50)	(0.50)	(0.50)			
Home Language English	0.28	0.20	0.38	-0.08	0.10	0.18**
	(0.45)	(0.40)	(0.49)			
Cognitive Reflection Score	2.07	2.02	2.05	-0.05	-0.02	0.03
	(0.94)	(0.93)	(1.05)			
N	57	65	60			

(i) Standard deviations in parentheses

(ii) T-test for difference being significant: * = 10%, ** = 5%, *** = 1%

Table 10: Average Updating Behavior across Treatments (Full Sample)

	T1: SYMMETRIC		T2: COMBINED		T3: SEPARATE	
	Reported (1a)	Corrected (1b)	Reported (2a)	Corrected (2b)	Reported (3a)	Corrected (3b)
OLS						
δ	0.74	0.73	0.80	0.80	0.84	0.86
	(0.05)***	(0.06)***	(0.03)***	(0.03)***	(0.04)***	(0.03)***
γ_a	0.95	0.92	0.86	0.91	1.00	0.93
	(0.10)	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)
γ_b	0.83	0.78	0.99	0.90	0.96	0.91
	(0.10)*	(0.10)**	(0.14)	(0.12)	(0.10)	(0.09)
p ($H_0 : \gamma_a = \gamma_b$)	0.26	0.19	0.25	0.90	0.64	0.80
N	1,875	1,875	1,850	1,850	1,825	1,825
R^2	0.52	0.51	0.62	0.64	0.67	0.70
IV						
δ	0.99	1.00	0.96	0.95	0.99	1.00
	(0.03)	(0.03)	(0.02)	(0.02)**	(0.02)	(0.02)
γ_a	0.87	0.84	0.87	0.83	0.95	0.88
	(0.09)	(0.09)*	(0.10)	(0.09)*	(0.11)	(0.09)
γ_b	0.82	0.78	0.89	0.89	0.92	0.88
	(0.09)*	(0.09)**	(0.12)	(0.12)	(0.09)	(0.08)
p ($H_0 : \gamma_a = \gamma_b$)	0.47	0.44	0.83	0.51	0.74	0.98
N	1,875	1,875	1,850	1,850	1,825	1,825
1st Stage F	70.89	82.63	82.83	82.45	54.72	59.90

(i) Standard errors in parentheses (clustered at the individual level).
(ii) All coefficients are significantly different from 0 at the 1% level. Therefore, t-tests of the null hypothesis (H_0: Coefficient = 1) are reported: * = 10%, ** = 5%, *** = 1%.
(iii) The rows corresponding to p ($H_0 : \gamma_a = \gamma_b$) report the p-statistic from a t-test of the equality of the coefficients γ_a and γ_b (i.e. a test of the asymmetric updating hypothesis).

Notes about relating the instructions to the discussion in the paper:

1. In the instructions below, the differences in the text between the treatments has been highlighted using a different colour for each of the three treatments – orange for T1: the SYMMETRIC treatment; purple for T2: the ASYMMETRIC COMBINED treatment; and green for T3: the ASYMMETRIC SEPARATE TREATMENT. For tables that differ across treatments, I have noted in the heading which is the relevant treatment.
2. Astute readers will notice that in the instructions Urn B corresponds to the state paying a larger bonus payment, while in the text of the paper state A is always the preferred state. This reversal of the labels A and B was made for expositional simplicity in the paper.

Stages of Today's Experiment

Today's experiment will have three stages. In each of the first two stages of the experiment, you can earn money. In addition to this, you will be paid a fixed fee for completing the third stage of £5, as well as a fixed participation fee of £5.

Within each stage, your earnings will depend partly on your decisions and partly on chance. The three stages are completely separate from one another – the choices made in one stage have no influence over the earnings from another stage. At the end of the third stage, your earnings will be calculated and you will be paid, privately. After this, the experiment will end.

Brief Overview of Stages of the Experiment:

<u>Stage 1:</u> In the first stage, you will face a task that involves estimating the likelihood of an event taking place. The more accurate your estimates, the higher your earnings will be on average. At the end of the experiment, one of these choices will be randomly selected and will determine your payment from Stage 1.

<u>Stage 2:</u> In the second stage, you will make a series of choices. At the end of the experiment, one of these choices will be randomly selected and will determine your payment from Stage 2.

<u>Stage 3:</u> You will be paid a fixed fee of £5 for completing this section. The answers to these questions will not affect your earnings, but are important for our study, so please try to answer them as accurately as possible.

After you have completed all three stages, we will calculate your earnings from Stage 1 and Stage 2 and add these to the £5 fixed fee from Stage 3, as well as the £5 participation fee. Therefore, you will earn at minimum £10 for completing the experiment and the remainder will depend on your choices in Stage 1 and Stage 2 as well as luck.

If you have any questions, please raise your hand. If not, we will proceed to Stage 1.

Instructions for Stage 1

Basic Setup

In this stage there will be five rounds. In each round, there are two possible "urns", each containing red and blue balls. In each round one of the two urns will be randomly chosen. You will not be told which urn has been chosen, but you will be given five pieces of information that will help you to decide which urn is more likely to be the one that was chosen in that round.

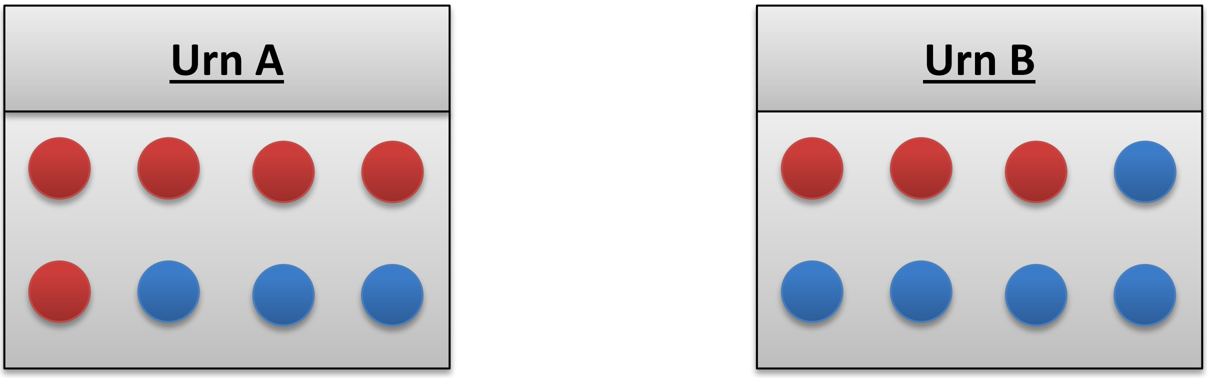
In each round you will be asked to make six probability judgements, expressed in percentages, about which urn was chosen, once before receiving any information and once after receiving each of these five pieces of information. Therefore, in Stage 1, you will make in total 30 probability judgements (6 probability judgements in each of 5 rounds).

Your payment for this round will be determined by randomly selecting one of these probability judgements in one of these rounds. While your payment will also depend on luck, the more accurate your chosen probability judgement, the higher your payoff will be on average.

Therefore, since the probability judgement that determines your payment will be chosen at random, you should pay attention and choose each of these probability judgements as accurately as possible if you would like to try to ensure that you receive a high payment. The specific details of how your payment will be determined are explained below.

Urn Selection

The two urns will be called Urn A and Urn B. Urn A contains 5 red balls and 3 blue balls; while Urn B contains 3 red balls and 5 blue balls, as depicted in the picture below.



In each round, one of these two urns, Urn A or Urn B, will be randomly chosen through the computer rolling a six sided die[1], however the chance of each urn being chosen will differ across rounds. For example, in one of the rounds Urn A will be used if the dice shows 1, 2, or 3 and Urn B being used if the die shows 4, 5, or 6. Therefore, in this round, each urn is *equally likely* to be used.

In another round, Urn A will be chosen if the die shows a 1, while Urn B will be chosen if the die shows 2, 3, 4, 5, or 6. In this round, Urn B is *five times more likely* to be chosen. You will be told at

[1] The word "die" is used as the singular for the word "dice" in this experiment.

the beginning of each round which numbers on the die will lead to Urn A being chosen and which will lead to Urn B being chosen. There will be five rounds in total. In each round, a new urn is chosen.

Information

In order to help you in making your probability judgements, in each round you will be provided with five pieces of information regarding which urn has been chosen.

This information will come in the form of a series of five ball draws from the chosen urn in that round. After each ball is drawn, you will observe the colour of the ball and then it will be replaced in the chosen urn. Therefore, within a specific round, the chosen urn will always contain the same 8 balls, and for each draw the computer will randomly select one of these balls with equal chance.

In each round, you will be asked to state six probability judgements, expressed in percentages, regarding the chance that Urn B is being used: once before you have seen any ball draws, and then once after each of the five draws.

Earnings and Experimental Points:

During experiment, you will earn experimental points. These points will be determined by the probability judgments you record and by whether the urn that is randomly chosen by the computer is actually Urn A or Urn B. The experimental points that you earn will be converted into real money at the end of the experiment. The rate of conversion is £1 = 6 000 points.

As discussed above, in each of the 5 rounds, you will report your probability judgement 6 times. Therefore, in total you will report your probability judgement about the chance of Urn B being used 30 times in Stage 1. At the very end of the experiment, one of these 30 reported probability judgements will be chosen at random to determine the points that you will earn from Stage 1. This will be done by first choosing one of the 5 rounds at random, and then within the chosen round, choosing one of the 6 reported probability judgements.

Your payment will then consist of two components:

(1) **A Probability Judgement Payment:** This payment depends on what you write down for the chosen probability judgement as well as the Urn that is used in the chosen round.
(2) **An Urn Bonus**: This is a bonus payment that depends on which Urn was being used in the round that is chosen for payment at the end of the experiment. If Urn A is chosen, this is 600 experimental points; if Urn B is chosen, this is [also 600 / 60 000 / 60 000] experimental points.

[We have combined these two payments / We have combined these two payments / The *probability judgement payment* is summarised] in the **'Stage 1: Score Sheet'** in front of you to show you what you would earn for each possible probability judgement that you might write down, between 0 and 100, for both the case where Urn A is the urn being used and for the case where Urn B is the urn being used. [_ / _ / In addition to this payment, you will receive the *urn bonus*, which will depend on which urn is being used.]

The payments in this table are designed in a way that makes it in your best interests to truthfully report your actual probability judgement if you want to ensure your payment is as high as possible, since the more accurate the probability judgement you report, the higher your payment will be on average.

If you look at your `Stage 1: Score Sheet', you will see that the higher the probability judgement you report about Urn B being the urn that is being used, the more points you will receive if Urn B is the chosen urn, and the fewer points you will receive if Urn A is the chosen urn.

Similarly, the lower the probability judgement you report about Urn B being used, the more points you will receive if Urn A is actually being used, and the fewer points you will receive if Urn B is actually being used.

How your score is calculated: A Summary

Your score for this stage consists of two components:

1. Firstly, one of the 6 recorded probability judgements that you write down in this round will be randomly chosen to determine the first part of your score.
2. Secondly, you will receive a bonus of [600 / 60 000/ 60 000] points if Urn B is the urn that is being used, and 600 points if Urn A is being used, in the round that is randomly chosen for payment.

[These two components are combined together / These two components are combined together / The first component, the *probability judgement payment*, is summarised] in the 'Stage 1: Score Sheet'

Procedure: A Summary

In each round you will proceed according to the following procedure:

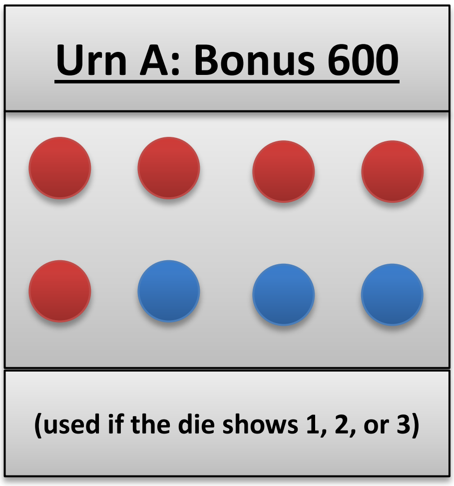
1. *Urn selection:* First, the computer will choose which Urn will be used through the throw of a six sided die. You will not observe this die throw and therefore will not know which Urn is selected.

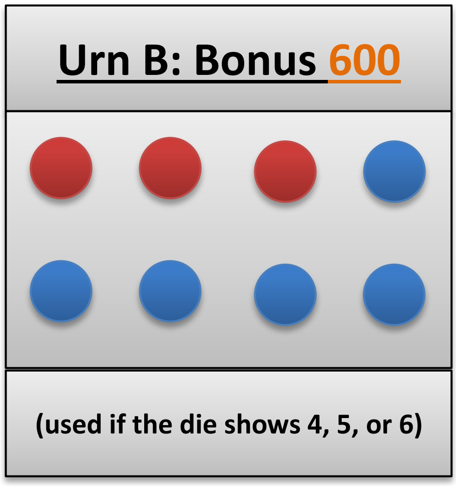
2. *Drawing balls*: Once the Urn has been selected, the computer will draw a sequence of 5 balls from the chosen Urn. *After each draw, the ball will be replaced in the container, so that each draw is made from the same 8 balls*. Therefore, it is possible, for example, to see the same exact ball drawn 5 times.

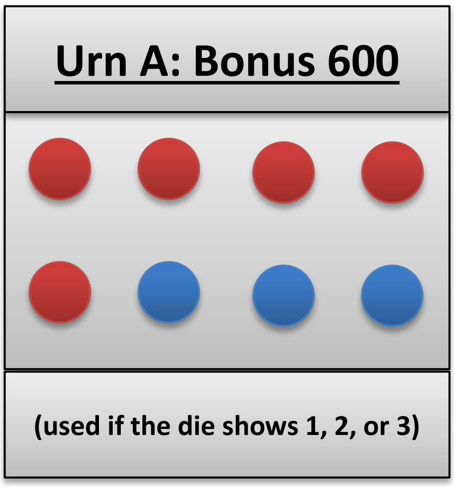
3. *Recording your probability judgements*: On the computer in front of you, you will be asked to record your probability judgements about the likelihood that Urn B is being used (similar to the example table below). You will then record your probability judgement, a number between 0 and 100 (up to two decimal places), about the likelihood that Urn B is being used. In other words, you should write down the percentage chance that you think Urn B is being used.

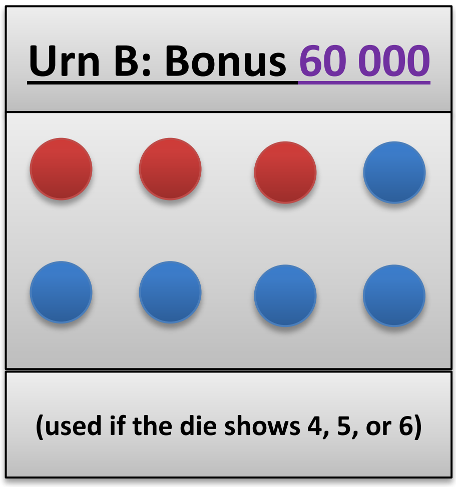
A Hypothetical Example:

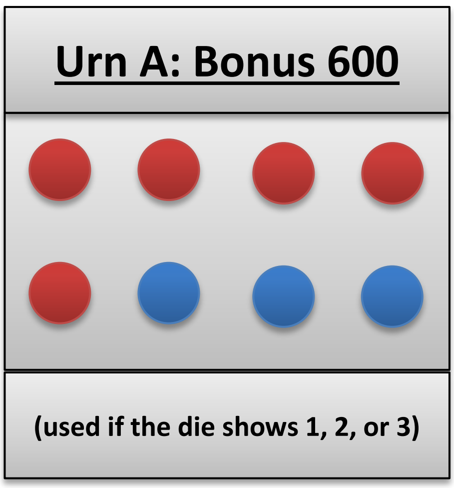
Just to illustrate how this process works, consider the following hypothetical example. Suppose that an imaginary person, Amy, is a participant in the experiment. In one of the rounds, Amy is told that Urn A will be chosen if the die shows 1, 2, or 3 and Urn B will be chosen if the die shows 4, 5, or 6.

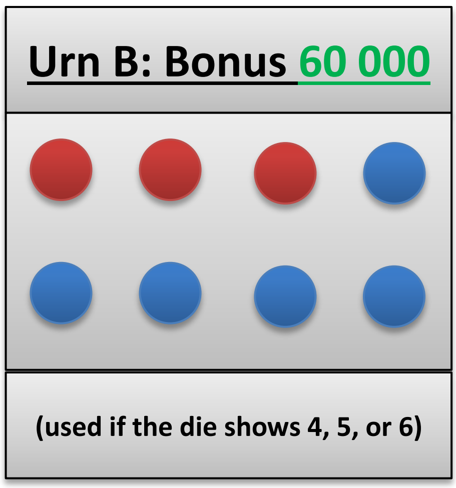












In order to show you how the payments work, suppose Amy wrote down the six probability judgements 80, 50, 20, 55, 85, 50 (as in the table below, although you might not think they were very good choices). Then if the fifth row (ball draw number 4) of this round is randomly selected for her payment and the actual urn being used is Urn B, then Amy would receive [12 330 points / 71 730 points / 11 730 points plus the bonus of 60 000, giving 71 730 points in total]. If the fifth row of this round is randomly selected for her payment and Urn A is being used, then Amy would receive [3 930 points / 3 930 points / 3 330 points plus the bonus of 600, giving 3 930 points in total].

Table 1: Example showing how Amy's score is determined for Stage 1 [T1: Symmetric]

			True Urn	
Ball Draw Number	**Ball Colour**	**Probability judgement about likelihood of Urn B**	**Urn A**	**Urn B**
	-	**80**	4 920	12 120
1	**R**	**50**	9 600	9 600
2	**R**	**20**	12 120	4 920
3	**R**	**55**	8 970	10 170
4	**R**	**85**	3 930	12 330
5	**B**	**50**	9 600	9 600

Table 2: Example showing how Amy's score is determined for Stage 1 [T2: Combined]

			True Urn	
Ball Draw Number	**Ball Colour**	**Probability judgement about likelihood of Urn B**	**Urn A**	**Urn B**
	-	**80**	4 920	71 520
1	**R**	**50**	9 600	69 000
2	**R**	**20**	12 120	64 320
3	**R**	**55**	8 970	69 570
4	**R**	**85**	3 930	71 730
5	**B**	**50**	9 600	69 000

Table 3: Example showing how Amy's score is determined for Stage 1 [T3: Separate]

			True Urn	
Ball Draw Number	**Ball Colour**	**Probability judgement about likelihood of Urn B**	**Urn A**	**Urn B**
	-	**80**	4 320	11 520
1	**R**	**50**	9 000	9 000
2	**R**	**20**	11 520	4 320
3	**R**	**55**	8 370	9 570
4	**R**	**85**	3 330	11 730
5	**B**	**50**	9 000	9 000

<u>Stage 1: Score Sheet</u> [T1: Symmetric]

Probability judgement for Urn B	True Urn			Probability judgement for Urn B	True Urn			Probability judgement for Urn B	True Urn	
	Urn A	Urn B			Urn A	Urn B			Urn A	Urn B
100	600	12 600		**66**	7 373	11 213		**32**	11 371	7 051
99	839	12 599		**65**	7 530	11 130		**31**	11 447	6 887
98	1 075	12 595		**64**	7 685	11 045		**30**	11 520	6 720
97	1 309	12 589		**63**	7 837	10 957		**29**	11 591	6 551
96	1 541	12 581		**62**	7 987	10 867		**28**	11 659	6 379
95	1 770	12 570		**61**	8 135	10 775		**27**	11 725	6 205
94	1 997	12 557		**60**	8 280	10 680		**26**	11 789	6 029
93	2 221	12 541		**59**	8 423	10 583		**25**	11 850	5 850
92	2 443	12 523		**58**	8 563	10 483		**24**	11 909	5 669
91	2 663	12 503		**57**	8 701	10 381		**23**	11 965	5 485
90	2 880	12 480		**56**	8 837	10 277		**22**	12 019	5 299
89	3 095	12 455		**55**	8 970	10 170		**21**	12 071	5 111
88	3 307	12 427		**54**	9 101	10 061		**20**	12 120	4 920
87	3 517	12 397		**53**	9 229	9 949		**19**	12 167	4 727
86	3 725	12 365		**52**	9 355	9 835		**18**	12 211	4 531
85	3 930	12 330		**51**	9 479	9 719		**17**	12 253	4 333
84	4 133	12 293		**50**	9 600	9 600		**16**	12 293	4 133
83	4 333	12 253		**49**	9 719	9 479		**15**	12 330	3 930
82	4 531	12 211		**48**	9 835	9 355		**14**	12 365	3 725
81	4 727	12 167		**47**	9 949	9 229		**13**	12 397	3 517
80	4 920	12 120		**46**	10 061	9 101		**12**	12 427	3 307
79	5 111	12 071		**45**	10 170	8 970		**11**	12 455	3 095
78	5 299	12 019		**44**	10 277	8 837		**10**	12 480	2 880
77	5 485	11 965		**43**	10 381	8 701		**9**	12 503	2 663
76	5 669	11 909		**42**	10 483	8 563		**8**	12 523	2 443
75	5 850	11 850		**41**	10 583	8 423		**7**	12 541	2 221
74	6 029	11 789		**40**	10 680	8 280		**6**	12 557	1 997
73	6 205	11 725		**39**	10 775	8 135		**5**	12 570	1 770
72	6 379	11 659		**38**	10 867	7 987		**4**	12 581	1 541
71	6 551	11 591		**37**	10 957	7 837		**3**	12 589	1 309
70	6 720	11 520		**36**	11 045	7 685		**2**	12 595	1 075
69	6 887	11 447		**35**	11 130	7 530		**1**	12 599	839
68	7 051	11 371		**34**	11 213	7 373		**0**	12 600	600
67	7 213	11 293		**33**	11 293	7 213				

Stage 1: Score Sheet [T2: Combined]

Probability judgement for Urn B	True Urn		Probability judgement for Urn B	True Urn		Probability judgement for Urn B	True Urn	
	Urn A	Urn B		Urn A	Urn B		Urn A	Urn B
100	600	72 000	**66**	7 373	70 613	**32**	11 371	66 451
99	839	71 999	**65**	7 530	70 530	**31**	11 447	66 287
98	1 075	71 995	**64**	7 685	70 445	**30**	11 520	66 120
97	1 309	71 989	**63**	7 837	70 357	**29**	11 591	65 951
96	1 541	71 981	**62**	7 987	70 267	**28**	11 659	65 779
95	1 770	71 970	**61**	8 135	70 175	**27**	11 725	65 605
94	1 997	71 957	**60**	8 280	70 080	**26**	11 789	65 429
93	2 221	71 941	**59**	8 423	69 983	**25**	11 850	65 250
92	2 443	71 923	**58**	8 563	69 883	**24**	11 909	65 069
91	2 663	71 903	**57**	8 701	69 781	**23**	11 965	64 885
90	2 880	71 880	**56**	8 837	69 677	**22**	12 019	64 699
89	3 095	71 855	**55**	8 970	69 570	**21**	12 071	64 511
88	3 307	71 827	**54**	9 101	69 461	**20**	12 120	64 320
87	3 517	71 797	**53**	9 229	69 349	**19**	12 167	64 127
86	3725	71 765	**52**	9 355	69 235	**18**	12 211	63 931
85	3 930	71 730	**51**	9 479	69 119	**17**	12 253	63 733
84	4 133	71 693	**50**	9 600	69 000	**16**	12 293	63 533
83	4 333	71 653	**49**	9 719	68 879	**15**	12 330	63 330
82	4 531	71 611	**48**	9 835	68 755	**14**	12 365	63 125
81	4 727	71 567	**47**	9 949	68 629	**13**	12 397	62 917
80	4 920	71 520	**46**	10 061	68 501	**12**	12 427	62 707
79	5 111	71 471	**45**	10 170	68 370	**11**	12 455	62 495
78	5 299	71 419	**44**	10 277	68 237	**10**	12 480	62 280
77	5485	71 365	**43**	10 381	68 101	**9**	12 503	62 063
76	5 669	71 309	**42**	10 483	67 963	**8**	12 523	61 843
75	5 850	71 250	**41**	10 583	67 823	**7**	12 541	61 621
74	6 029	71 189	**40**	10 680	67 680	**6**	12 557	61 397
73	6 205	71 125	**39**	10 775	67 535	**5**	12 570	61 170
72	6 379	71 059	**38**	10 867	67 387	**4**	12 581	60 941
71	6 551	70 991	**37**	10 957	67 237	**3**	12 589	60 709
70	6 720	70 920	**36**	11 045	67 085	**2**	12 595	60 475
69	6 887	70 847	**35**	11 130	66 930	**1**	12 599	60 239
68	7 051	70 771	**34**	11 213	66 773	**0**	12 600	60 000
67	7 213	70 693	**33**	11 293	66 613			

Stage 1: Score Sheet [T3: Separate]

Probability judgement for Urn B	True Urn			Probability judgement for Urn B	True Urn			Probability judgement for Urn B	True Urn	
	Urn A	Urn B			Urn A	Urn B			Urn A	Urn B
100	0	12 000		**66**	6 773	10 613		**32**	10 771	6 451
99	239	11 999		**65**	6 930	10 530		**31**	10 847	6 287
98	475	11 995		**64**	7 085	10 445		**30**	10 920	6 120
97	709	11 989		**63**	7 237	10 357		**29**	10 991	5 951
96	941	11 981		**62**	7 387	10 267		**28**	11 059	5 779
95	1 170	11 970		**61**	7 535	10 175		**27**	11 125	5 605
94	1 397	11 957		**60**	7 680	10 080		**26**	11 189	5 429
93	1 621	11 941		**59**	7 823	9 983		**25**	11 250	5 250
92	1 843	11 923		**58**	7 963	9 883		**24**	11 309	5 069
91	2 063	11 903		**57**	8 101	9 781		**23**	11 365	4 885
90	2 280	11 880		**56**	8 237	9 677		**22**	11 419	4 699
89	2 495	11 855		**55**	8 370	9 570		**21**	11 471	4 511
88	2 707	11 827		**54**	8 501	9 461		**20**	11 520	4 320
87	2 917	11 797		**53**	8 629	9 349		**19**	11 567	4 127
86	3 125	11 765		**52**	8 755	9 235		**18**	11 611	3 931
85	3 330	11 730		**51**	8 879	9 119		**17**	11 653	3 733
84	3 533	11 693		**50**	9 000	9 000		**16**	11 693	3 533
83	3 733	11 653		**49**	9 119	8 879		**15**	11 730	3 330
82	3 931	11 611		**48**	9 235	8 755		**14**	11 765	3 125
81	4 127	11 567		**47**	9 349	8 629		**13**	11 797	2 917
80	4 320	11 520		**46**	9 461	8 501		**12**	11 827	2 707
79	4 511	11 471		**45**	9 570	8 370		**11**	11 855	2 495
78	4 699	11 419		**44**	9 677	8 237		**10**	11 880	2 280
77	4 885	11 365		**43**	9 781	8 101		**9**	11 903	2 063
76	5 069	11 309		**42**	9 883	7 963		**8**	11 923	1 843
75	5 250	11 250		**41**	9 983	7 823		**7**	11 941	1 621
74	5 429	11 189		**40**	10 080	7 680		**6**	11 957	1 397
73	5 605	11 125		**39**	10 175	7 535		**5**	11 970	1 170
72	5 779	11 059		**38**	10 267	7 387		**4**	11 981	941
71	5 951	10 991		**37**	10 357	7 237		**3**	11 989	709
70	6 120	10 920		**36**	10 445	7 085		**2**	11 995	475
69	6 287	10 847		**35**	10 530	6 930		**1**	11 999	239
68	6 451	10 771		**34**	10 613	6 773		**0**	12 000	0
67	6 613	10 693		**33**	10 693	6 613				

Stage 2: Part A

Part A consists of 20 statements. As in Stage 1, you will be asked to report a probability judgement for the likelihood of an event occurring. The difference here is that you will be making probability judgements for the truth of *statements* that depend on the computer randomly choosing a number between 1 and 100. Each number between 1 and 100 has an equal probability of being chosen.

An example of a possible **statement** in Stage 2 is:

"the number the computer chooses will be between 1 and 75".

This statement will be evaluated as true if the randomly chosen number is a value between 1 and 75 (including 1 and 75). This statement will be evaluated as false if the computer randomly chooses a number that is higher than 75. So, for example, if the computer randomly chooses 91 then the statement would be evaluated as false. However, if the computer were to randomly choose the number 71, then the statement would be evaluated as true.

During Stage 2: Part A, the computer will show you 20 statements of this type. For each of these statements, you should write down a probability judgement as you did in Part A of the experiment.

Recall that your earnings from Stage 2 will be determined by randomly selecting one of your decisions to determine your payment. Therefore, you should make your decisions carefully and for each decision act as if that is the one that will determine your payment for Stage 2. If this chosen decision is one of the probability judgements from Part A then the computer will randomly choose a number between 1 and 100, with equal probability, to evaluate this statement and determine your earnings from Stage 2.

Also, in this section there is no right or wrong answer; you can choose what you want best. You should use the **'Stage 2: Part A: Score Sheet'** to see what your payment will be if the statement is true and if it is false, for each probability judgement you could write down. [_ / _ / In addition to the payments in the 'Stage 2: Part A: Score Sheet', if one of the probability judgements from Part A is chosen to determine your Stage 2 payment, then you will receive a **bonus** of 60 000 if the chosen statement is evaluated as true, and a bonus of 600 if the chosen statement is evaluated as false.]

A Hypothetical Example

Just to illustrate how this process works, suppose that our imaginary person, Amy, is choosing what probability judgement to write down for the statement "the number the computer chooses will be between 1 and 75". Suppose that Amy writes down 10. Then Amy would receive [12 480 points / 12 480 points / 11 880 plus a bonus of 600, giving a total of 12 480 points] if the random number chosen by the computer is between 76 and 100 and she would receive [2 880 points / 62 280 points / 2 280 plus a bonus of 60 000, giving 62 280 points] if the randomly chosen number is between 1 and 75.

If instead, Amy had written down 50, then she would receive [9 600 points / 9 600 points / 9 000 plus a bonus of 600, giving 9 600 points] if the random number is between 76 and 100; and [9 600 points / 69 000 points / 9 000 plus a bonus of 60 000, giving 69 000 points] if the random number is between 1 and 75.

Table 4: Example showing how Amy's score is determined for Stage 2: Part A [T1: Symmetric]

Probability judgement for the statement being 'True'	Statement	
	False	True
10	12 480	2 880
50	9 600	9 600
100	600	12 600

Table 5: Example showing how Amy's score is determined for Stage 2: Part A [T2: Symmetric]

Probability judgement for the statement being 'True'	Statement	
	False	True
10	12 480	62 280
50	9 600	69 000
100	600	72 000

Table 6: Example showing how Amy's score is determined for Stage 2: Part A [T3: Separate]

Probability judgement for the statement being 'True'	Statement	
	False	True
10	11 880	2 280
50	9 000	9 000
100	0	12 000

We will now proceed to carrying out Stage 2: Part A of the experiment. Before we do, if you have any questions at this moment, raise your hand. The experimenter will come to you.

Stage 2: Part A: Score Sheet [T1: Symmetric]

Probability judgement for the statement being 'True'	Statement		Probability judgement for the statement being 'True'	Statement		Probability judgement for the statement being 'True'	Statement	
	False	True		False	True		False	True
100	600	12 600	**66**	7 373	11 213	**32**	11 371	7 051
99	839	12 599	**65**	7 530	11 130	**31**	11 447	6 887
98	1 075	12 595	**64**	7 685	11 045	**30**	11 520	6 720
97	1 309	12 589	**63**	7 837	10 957	**29**	11 591	6 551
96	1 541	12 581	**62**	7 987	10 867	**28**	11 659	6 379
95	1 770	12 570	**61**	8 135	10 775	**27**	11 725	6 205
94	1 997	12 557	**60**	8 280	10 680	**26**	11 789	6 029
93	2 221	12 541	**59**	8 423	10 583	**25**	11 850	5 850
92	2 443	12 523	**58**	8 563	10 483	**24**	11 909	5 669
91	2 663	12 503	**57**	8 701	10 381	**23**	11 965	5 485
90	2 880	12 480	**56**	8 837	10 277	**22**	12 019	5 299
89	3 095	12 455	**55**	8 970	10 170	**21**	12 071	5 111
88	3 307	12 427	**54**	9 101	10 061	**20**	12 120	4 920
87	3 517	12 397	**53**	9 229	9 949	**19**	12 167	4 727
86	3 725	12 365	**52**	9 355	9 835	**18**	12 211	4 531
85	3 930	12 330	**51**	9 479	9 719	**17**	12 253	4 333
84	4 133	12 293	**50**	9 600	9 600	**16**	12 293	4 133
83	4 333	12 253	**49**	9 719	9 479	**15**	12 330	3 930
82	4 531	12 211	**48**	9 835	9 355	**14**	12 365	3 725
81	4 727	12 167	**47**	9 949	9 229	**13**	12 397	3 517
80	4 920	12 120	**46**	10 061	9 101	**12**	12 427	3 307
79	5 111	12 071	**45**	10 170	8 970	**11**	12 455	3 095
78	5 299	12 019	**44**	10 277	8 837	**10**	12 480	2 880
77	5 485	11 965	**43**	10 381	8 701	**9**	12 503	2 663
76	5 669	11 909	**42**	10 483	8 563	**8**	12 523	2 443
75	5 850	11 850	**41**	10 583	8 423	**7**	12 541	2 221
74	6 029	11 789	**40**	10 680	8 280	**6**	12 557	1 997
73	6 205	11 725	**39**	10 775	8 135	**5**	12 570	1 770
72	6 379	11 659	**38**	10 867	7 987	**4**	12 581	1 541
71	6 551	11 591	**37**	10 957	7 837	**3**	12 589	1 309
70	6 720	11 520	**36**	11 045	7 685	**2**	12 595	1 075
69	6 887	11 447	**35**	11 130	7 530	**1**	12 599	839
68	7 051	11 371	**34**	11 213	7 373	**0**	12 600	600
67	7 213	11 293	**33**	11 293	7 213			

<u>Stage 2: Part A: Score Sheet</u> [T2: Combined]

Probability judgement for the statement being ‘True’	Statement		Probability judgement for the statement being ‘True’	Statement		Probability judgement for the statement being ‘True’	Statement	
	False	True		False	True		False	True
100	600	72 000	**66**	7 373	70 613	**32**	11 371	66 451
99	839	71 999	**65**	7 530	70 530	**31**	11 447	66 287
98	1 075	71 995	**64**	7 685	70 445	**30**	11 520	66 120
97	1 309	71 989	**63**	7 837	70 357	**29**	11 591	65 951
96	1 541	71 981	**62**	7 987	70 267	**28**	11 659	65 779
95	1 770	71 970	**61**	8 135	70 175	**27**	11 725	65 605
94	1 997	71 957	**60**	8 280	70 080	**26**	11 789	65 429
93	2 221	71 941	**59**	8 423	69 983	**25**	11 850	65 250
92	2 443	71 923	**58**	8 563	69 883	**24**	11 909	65 069
91	2 663	71 903	**57**	8 701	69 781	**23**	11 965	64 885
90	2 880	71 880	**56**	8 837	69 677	**22**	12 019	64 699
89	3 095	71 855	**55**	8 970	69 570	**21**	12 071	64 511
88	3 307	71 827	**54**	9 101	69 461	**20**	12 120	64 320
87	3 517	71 797	**53**	9 229	69 349	**19**	12 167	64 127
86	3725	71 765	**52**	9 355	69 235	**18**	12 211	63 931
85	3 930	71 730	**51**	9 479	69 119	**17**	12 253	63 733
84	4 133	71 693	**50**	9 600	69 000	**16**	12 293	63 533
83	4 333	71 653	**49**	9 719	68 879	**15**	12 330	63 330
82	4 531	71 611	**48**	9 835	68 755	**14**	12 365	63 125
81	4 727	71 567	**47**	9 949	68 629	**13**	12 397	62 917
80	4 920	71 520	**46**	10 061	68 501	**12**	12 427	62 707
79	5 111	71 471	**45**	10 170	68 370	**11**	12 455	62 495
78	5 299	71 419	**44**	10 277	68 237	**10**	12 480	62 280
77	5485	71 365	**43**	10 381	68 101	**9**	12 503	62 063
76	5 669	71 309	**42**	10 483	67 963	**8**	12 523	61 843
75	5 850	71 250	**41**	10 583	67 823	**7**	12 541	61 621
74	6 029	71 189	**40**	10 680	67 680	**6**	12 557	61 397
73	6 205	71 125	**39**	10 775	67 535	**5**	12 570	61 170
72	6 379	71 059	**38**	10 867	67 387	**4**	12 581	60 941
71	6 551	70 991	**37**	10 957	67 237	**3**	12 589	60 709
70	6 720	70 920	**36**	11 045	67 085	**2**	12 595	60 475
69	6 887	70 847	**35**	11 130	66 930	**1**	12 599	60 239
68	7 051	70 771	**34**	11 213	66 773	**0**	12 600	60 000
67	7 213	70 693	**33**	11 293	66 613			

Stage 2: Part A: Score Sheet [T3: Separate]

Probability judgement for the statement being 'True'	Statement			Probability judgement for the statement being 'True'	Statement			Probability judgement for the statement being 'True'	Statement	
	False	True			False	True			False	True
100	0	12 000		**66**	6 773	10 613		**32**	10 771	6 451
99	239	11 999		**65**	6 930	10 530		**31**	10 847	6 287
98	475	11 995		**64**	7 085	10 445		**30**	10 920	6 120
97	709	11 989		**63**	7 237	10 357		**29**	10 991	5 951
96	941	11 981		**62**	7 387	10 267		**28**	11 059	5 779
95	1 170	11 970		**61**	7 535	10 175		**27**	11 125	5 605
94	1 397	11 957		**60**	7 680	10 080		**26**	11 189	5 429
93	1 621	11 941		**59**	7 823	9 983		**25**	11 250	5 250
92	1 843	11 923		**58**	7 963	9 883		**24**	11 309	5 069
91	2 063	11 903		**57**	8 101	9 781		**23**	11 365	4 885
90	2 280	11 880		**56**	8 237	9 677		**22**	11 419	4 699
89	2 495	11 855		**55**	8 370	9 570		**21**	11 471	4 511
88	2 707	11 827		**54**	8 501	9 461		**20**	11 520	4 320
87	2 917	11 797		**53**	8 629	9 349		**19**	11 567	4 127
86	3 125	11 765		**52**	8 755	9 235		**18**	11 611	3 931
85	3 330	11 730		**51**	8 879	9 119		**17**	11 653	3 733
84	3 533	11 693		**50**	9 000	9 000		**16**	11 693	3 533
83	3 733	11 653		**49**	9 119	8 879		**15**	11 730	3 330
82	3 931	11 611		**48**	9 235	8 755		**14**	11 765	3 125
81	4 127	11 567		**47**	9 349	8 629		**13**	11 797	2 917
80	4 320	11 520		**46**	9 461	8 501		**12**	11 827	2 707
79	4 511	11 471		**45**	9 570	8 370		**11**	11 855	2 495
78	4 699	11 419		**44**	9 677	8 237		**10**	11 880	2 280
77	4 885	11 365		**43**	9 781	8 101		**9**	11 903	2 063
76	5 069	11 309		**42**	9 883	7 963		**8**	11 923	1 843
75	5 250	11 250		**41**	9 983	7 823		**7**	11 941	1 621
74	5 429	11 189		**40**	10 080	7 680		**6**	11 957	1 397
73	5 605	11 125		**39**	10 175	7 535		**5**	11 970	1 170
72	5 779	11 059		**38**	10 267	7 387		**4**	11 981	941
71	5 951	10 991		**37**	10 357	7 237		**3**	11 989	709
70	6 120	10 920		**36**	10 445	7 085		**2**	11 995	475
69	6 287	10 847		**35**	10 530	6 930		**1**	11 999	239
68	6 451	10 771		**34**	10 613	6 773		**0**	12 000	0
67	6 613	10 693		**33**	10 693	6 613				