Online Appendix for Good News and Bad News are Still News: Experimental Evidence on Belief Updating

Alexander Coutts
Nova School of Business and Economics

April 3, 2018

Online Appendix

A Hedge Proof Design Details

Payoffs are determined in the following way, also described in Figure A1, and earlier utilized by Blanco et al. (2010).[1] In order to ensure that individuals have no incentive to hedge their probability reports, the world is partitioned into two disjoint states, the accuracy state and the prize state.[2]

With probability 0.5 the individual is paid solely according to her reported belief $\tilde{\pi}$ about whether event E occurred using the incentive compatible lottery method to elicit beliefs with an accuracy payment of $a > 0$ (accuracy state).

In the other state occurring with probability 0.5, the individual receives a guaranteed payment $\bar{a} \geq a$[3] and receives an additional \$80 if E occurs, but receives nothing extra if E does not occur (prize state). Her report of $\tilde{\pi}$ is no longer relevant in this prize state.

To be clear, two types of hedging are of concern in this experiment. The first is is hedging within the accuracy state, which is solved through use of the lottery method. The second is hedging across accuracy and prize states, which is solved through partitioning. In isolation, the lottery method is incentive compatible under the relatively weak assumption of probabilistic sophistication. However, the experiment design introduces further elements of randomization through the partitioning of the accuracy and prize states, *and* through randomly selection one decision for payment.

[1] It was also independently suggested to me by Christopher Woolnough, who I credit for the design in this paper.

[2] Hedging will be present whenever utility is not linear, for example with a concave utility function and a positive stake in an event an individual would prefer to report a lower than truthful $\tilde{\pi}$, since this will smooth consumption over the different states of the world. Karni and Safra (1995) show that without this partition, no elicitation procedure exists that induces truthful reporting, a fact that is sometimes overlooked in the experimental literature; see Armantier and Treich (2013).

[3] The payment of $\bar{a}$ is to ensure that the prize state is preferred to the accuracy state, required for an earlier theoretical extension; it is not necessary for any of the analysis.

As Blanco et al. (2010) note, partitioning the world into an accuracy and a prize state is akin to the standard procedure of introducing a new lottery and randomly selecting one lottery for payment.[4] Thus incentive compatibility in the broader experiment design holds for the class of preferences where payment is made by random selection of one task (or lottery). This is true when one assumes a statewise monotonicity condition, see Azrieli et al. (2018). This condition is equivalent to saying that subjects never choose dominated gambles, independent of other states.[5]

Two further issues on incentive compatibility deserve some mention here. First, with financial incentives it is possible to disentangle accuracy and prize payments. However, if subjects gain utility from beliefs about their ability, evidently the experiment is not able to create an analogous partition. Thus there may be distortions in elicited beliefs about performance on the quiz - note however that most of the results in this paper do not hinge on the inclusion of the quiz (Self) event.

Second, there is a potential concern which arises from paying only for the accuracy *or* the prize state, but never both. The implication is that a subject in the experiment knows with certainty that whenever her belief report is relevant, she will not have an opportunity to win the prize. Or vice-versa, whenever she has a chance to win the prize, her belief report is not relevant.[6]

In this case, the procedure would correctly capture the subject's belief about the event occurring in the event the prize is irrelevant, but the counterfactual belief would not be observed. If such belief patterns are occurring then it remains possible that subjects may hold biased beliefs, but the experiment is not designed to capture them. Under the assumption of monotonicity above, this does not cause an issue as subjects are assumed to form consistent beliefs about an event, which do not depend on the state.[7]

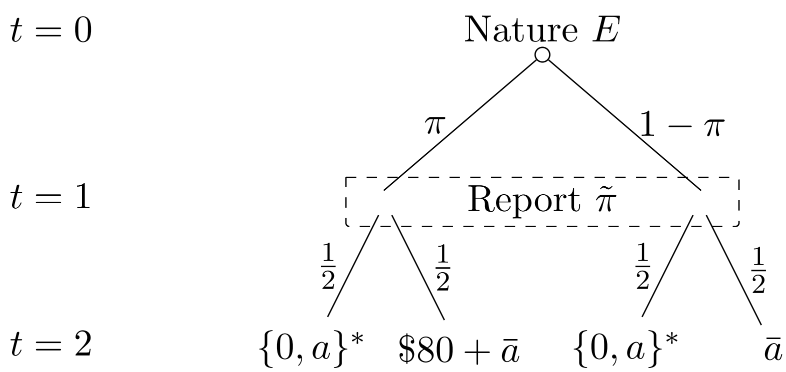
[4] In this case the other lottery is degenerate, as the individual does not make an active choice, but simply has the opportunity to receive a payment.

[5] The assumption of monotonicity is not completely innocuous. Assuming further that subjects reduce compound lotteries, it implies that subject's preferences must conform with expected utility, as detailed in Azrieli et al. (2018).

[6] I thank an anonymous referee for bringing this concern to my attention.

[7] Barron (2016) elicits beliefs about events with financial stakes without separating prize and accuracy payments (but addressing the hedging problem retroactively, using the "truth serum" of Offerman et al. (2009)). He does not find evidence of differential updating patterns with financial stakes, which suggests that this may not be a concern.

Figure A1: Illustration of Hedge Proof Design



*In the accuracy state the payoff is either 0 or a, depending on the reported belief $\tilde{\pi}$ and whether E occurred, according to the lottery method.

Nature determines outcome of binary event E. Individual submits report $\tilde{\pi}$ without knowing outcome of E, and payoff is determined according to the lottery method elicitation procedure.

B Updating Framework: By Event/Stake/Accuracy Payment

Here I replicate the primary analysis found in Table 1, looking at each of the financial stake and accuracy payment conditions separately. As can be seen in Table B1, there is no clear pattern that emerges within either the accuracy payment or within the financial stake conditions respectively. A formal statistical test confirms that I cannot reject equality between the \$0 and \$80 financial stake conditions, nor between the \$3, \$10, and \$20 accuracy payment conditions. This analysis suggests that different payments for accuracy do not alter updating behavior. Similarly, holding a large financial stake in an event does not alter updating behavior relative to holding no stake.

Table B1: Updating Beliefs for All Events: By Accuracy Payment and Financial Stake

Dependent Variable: Logit Posterior Belief						
	(1)	(2)	(3)	(4)	(5)	(6)
Regressor	Stake = 0	Stake = 80	Acc = 3	Acc = 10	Acc = 20	Total
δ	0.910***	0.918***	0.920***	0.922***	0.898***	0.914***
	(0.012)	(0.014)	(0.017)	(0.014)	(0.016)	(0.009)
β_1	0.587***	0.588***	0.560***	0.662***	0.540***	0.588***
	(0.045)	(0.043)	(0.054)	(0.063)	(0.059)	(0.034)
β_0	0.807***	0.780***	0.774***	0.749***	0.861*	0.793***
	(0.047)	(0.047)	(0.066)	(0.060)	(0.074)	(0.038)
P-Value ($\delta = 1$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P-Value ($\beta_0 = 1$)	0.0001	0.0000	0.0008	0.0001	0.0616	0.0000
Diff ($\beta_1 - \beta_0$)	−0.220	−0.192	−0.214	−0.086	−0.321	−0.205
P-Value ($\beta_1 = \beta_0$)	0.0001	0.0011	0.0042	0.2112	0.0000	0.0000
R^2	0.83	0.84	0.84	0.84	0.82	0.84
Observations	1704	1656	1128	1143	1089	3360

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

In Table B2 I replicate the primary analysis found in Table 2, but excluding any observations where an individual had a financial stake of $80. There do not appear to be any consistent differences in this subsample. In the final column of Table B2 I use the same sampling procedure as Mobius et al. (2014), in order to provide a more comparable estimation to their study for the Quiz event. One can see that the sampling procedure does not significantly alter the pattern of observed results.

Table B2: Updating Beliefs Within Events: No Financial Stake Only

	Dependent Variable: Logit Posterior Belief					
	(1)	(2)	(3)	(4)	(5)	(6)
Regressor	Easy Dice	Hard Dice	Weather	Quiz (S)	Quiz (O)	Quiz (M. et al)
δ	0.839***	0.897***	0.909***	0.924**	0.894*	0.918**
	(0.049)	(0.028)	(0.027)	(0.030)	(0.055)	(0.035)
β_1	0.317***	0.430***	0.683***	0.616***	0.816	0.714***
	(0.146)	(0.092)	(0.089)	(0.078)	(0.200)	(0.090)
β_0	1.073	0.815*	0.783***	0.799**	0.778	0.917
	(0.171)	(0.109)	(0.067)	(0.087)	(0.176)	(0.099)
P-Value ($\delta = 1$)	0.0013	0.0004	0.0009	0.0138	0.0603	0.0226
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0005	0.0000	0.3623	0.0021
P-Value ($\beta_0 = 1$)	0.6711	0.0906	0.0014	0.0224	0.2144	0.4049
Diff ($\beta_1 - \beta_0$)	−0.755	−0.384	−0.100	−0.184	0.038	−0.204
P-Value ($\beta_1 = \beta_0$)	0.0095	0.0136	0.3791	0.0472	0.9040	0.0521
R^2	0.66	0.77	0.73	0.83	0.79	0.84
Observations	435	421	447	294	107	225

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant. Includes only updated beliefs about events where individuals did not hold any additional financial stake in the outcome.

C Additional Tests of Bayes' Rule: Invariance, Sufficiency, and Stability

In this section I investigate three additional properties that are satisfied when updated beliefs follow Bayes' rule. First, the structure of Bayes' rule implies a sufficiency condition, that priors are sufficient statistics for all the information contained in past signals. In other words, after controlling for prior beliefs, lagged information does not significantly predict posterior beliefs. To examine whether updating behavior can be shown to satisfy the sufficiency condition I follow Mobius et al. (2014) and include lagged signals as independent variables. Table C1 shows the regressions that include these lagged signals, using only actively revised beliefs. There is some evidence that overall, the updating process may not satisfy the sufficiency condition, as the first signal received has a significant effect on belief updating in round 3.[8]

Table C1: Examining Sufficiency

Dependent Variable: Logit Posterior Belief		
	(1)	(2)
Regressor	Round 2	Round 3
δ	0.890***	0.880***
	(0.027)	(0.023)
β_1	1.030***	1.247***
	(0.065)	(0.074)
β_0	1.287***	1.347***
	(0.064)	(0.066)
β_{t-1}	0.052	0.048
	(0.045)	(0.042)
β_{t-2}		0.164***
		(0.042)
R^2	0.82	0.82
Observations	640	670

Analysis uses OLS regression. Difference is *significant from zero* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant. The sample is restricted to include only subjects who actively revised their beliefs in the direction predicted by Bayes' rule. β_{t-k} refers to the k^{th} lagged signal.

The next property Bayes' rule satisfies is stability: that updating remains stable across time. Looking across the three updating rounds in Table C2, there appear to be differences. Overall, I can reject equality across rounds 1 to 3 for δ, β_1, β_0 at conventional levels. This provides some evidence that updating is not stable across rounds. Finally, the invariance property is said to hold when $\delta = 1$, that is the change in logit beliefs depends only on past signals. $\delta = 1$ is rejected in the data at the 1% level. However, despite these three conditions not being met in the data, it is important to note that the magnitude of these deviations is reasonably small, in the sense that the resulting posteriors are very close to

[8] While Mobius et al. (2014) do not reject sufficiency, it is worth noting that the ratio of the values of coefficients on lagged signals to current signals is of the same magnitude.

Table C2: Examining Stability

Dependent Variable: Logit Posterior Belief				
Regressor	Round 1	Round 2	Round 3	All Rounds
δ	0.884***	0.926***	0.935***	0.914***
	(0.014)	(0.017)	(0.016)	(0.009)
β_1	0.468***	0.537***	0.800***	0.588***
	(0.047)	(0.046)	(0.062)	(0.034)
β_0	0.687***	0.788***	0.914	0.793***
	(0.045)	(0.053)	(0.055)	(0.038)
P-Value ($\delta = 1$)	0.0000	0.0000	0.0001	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0015	0.0000
P-Value ($\beta_0 = 1$)	0.0000	0.0001	0.1205	0.0000
Diff ($\beta_1 - \beta_0$)	−0.219	−0.250	−0.114	−0.205
P-Value ($\beta_1 = \beta_0$)	0.0002	0.0003	0.1592	0.0000
R^2	0.85	0.84	0.83	0.84
Observations	1180	1135	1045	3360
P-Value [Chow-test] for δ (Rounds 1-3)				0.0260
P-Value [Chow-test] for β_1 (Rounds 1-3)				0.0000
P-Value [Chow-test] for β_0 (Rounds 1-3)				0.0005

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

their Bayesian counterparts.[9]

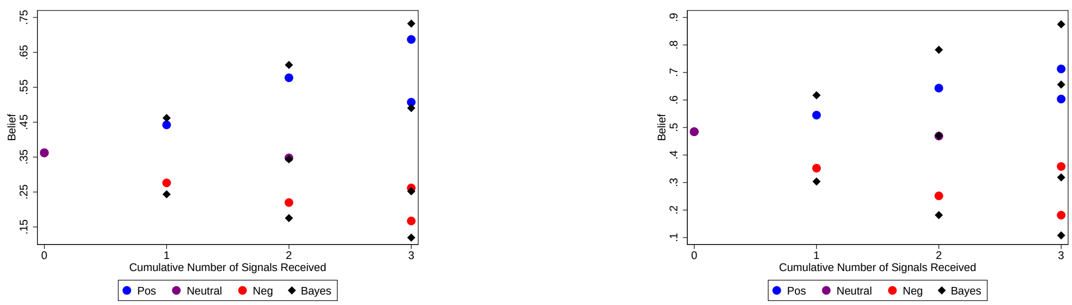
D Updating by Sequence of Signals Observed

Figure D1 presents an aggregate view of asymmetry, by plotting average posteriors in response to different sequences of observed signals, for both the aggregate data, and for moderate priors between 0.4 and 0.6. One can note that the asymmetry in the framework,

[9] As in Mobius et al. (2014) a concern is that β_1 and β_0 are functions of prior beliefs, but that effects cancel out to give a coefficient of δ closer to 1. To examine if this is a potential issue I check whether there are significant interaction effects between receiving affirmative signals, and the prior. These interactions are never significant at any reasonable significance level, indicating that this is not a problem for the data.

observed in Table 1, is not visibly present in the aggregate data. The reason is that the weight on the log odds ratio of prior beliefs is not unity. $\delta < 1$ manifests itself as over-weighting of probabilities for priors < 0.5, the majority of the data of this experiment. This masks the asymmetry in the framework, since it results in an upward shift of posterior beliefs, independent of the types of signals received.[10]

Figure D1: Updating in Response to Observed Signals



(a) All Data

(b) 40% < Prior < 60%

Average belief update following a sequence of cumulative signals (numbered on the horizontal axis), distinguishing cumulative signals in the positive direction (blue) from negative (red), as well as neutral (purple). For example, when the number of cumulative signals is 2, the possibilities are that a subject received 2 positive signals, 1 positive 1 negative, or 2 negative signals. The Bayesian benchmark is indicated by a black diamond.

E Robustness Checks

E.1 Different Values of the Prior

Of interest is to what extent the results in the paper could be explained by the fact that priors are on average lower than one-half.

Figure E1 presents the evolution of beliefs for different values of the prior (first reported beliefs). Updating appears conservative for low values of the prior, well calibrated for moderate values, and too responsive for high values of the prior. These patterns are suggestive that some of the differences in updating observed across events are driven by differences in average values of the prior, rather than differences in the events themselves. In particular, elicited priors for the two dice events and the quiz (other) event are significantly

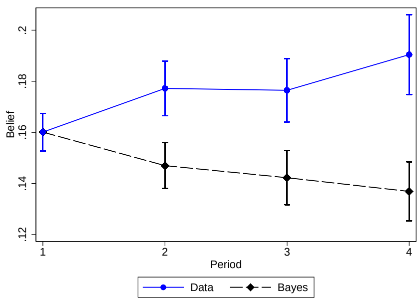
[10] There is some visible asymmetry when restricting priors to be moderate, between 0.4 and 0.6. This is intuitive, as the distortionary weighting of δ is weakest around 0.5, and hence posteriors are more closely matching the response to signals observed in the framework.

lower than those for the weather event and the quiz (self) event, and additionally exhibit substantially greater levels of conservatism.

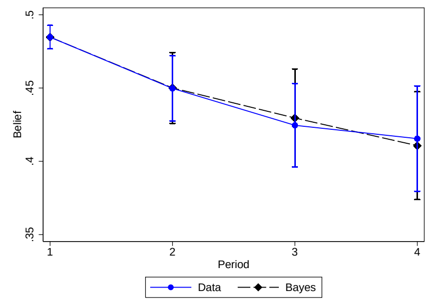
Table E1 examines the primary results of Table 1, but restricting priors to lie between 0.4 and 0.6. From the table, one is able to see that the results are very similar. Negative signals continue to be weighted significantly more than positive signals, and one cannot reject that the difference in any of the parameters in the valenced contexts (good/bad news) are the same as those found in the neutral (just news) contexts. Analogously, Table E2 presents the results of Table 2 with the same prior restrictions, finding no differences in the patterns of updating.

Table E3 splits the sample into priors less than one half or greater than one-half. There, one can see similar asymmetries across the two subsamples. One finding of note is that the value of δ is very close to one in Column (2), when priors are greater than one-half. In fact, this has implications for interpreting differences between patterns in the empirical framework and the raw data. It implies that an over-weighting of priors less than one half occurs (since $\delta < 1$ in this case), but no corresponding under-weighting of priors greater than one-half is occurring (since δ is approximately 1 in this case).

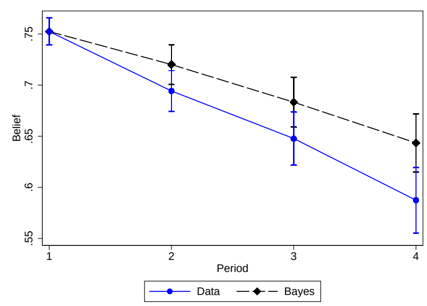
Figure E1: Evolution of Beliefs By Value of the Prior



(a) Prior $\leq$ 40%



(b) 40% $\leq$ Prior $\leq$ 60%



(c) 60% $\leq$ Prior

All events and all stake conditions, sample restricted to individuals updating with indicated prior. The path of beliefs starting from the prior (period 1), and after each sequential signal (periods 2 through 4). Average individual responses are the blue solid line, the Bayesian benchmark is marked as the black dashed line. Error bands represent 95% confidence intervals. $N = \{798, 185, 297\}$ average per round, respectively for (a)-(c).

Table E1: Updating Beliefs for All Events: 40% $\leq$ Prior $\leq$ 60%

Dependent Variable: Logit Posterior Belief			
Regressor	(1) Good/Bad News	(2) Just News	(3) All
δ^V	0.897		
	(0.091)		
β_1^V	0.586***		
	(0.085)		
β_0^V	0.777***		
	(0.078)		
δ^N		0.918	
		(0.055)	
β_1^N		0.576***	
		(0.102)	
β_0^N		0.755**	
		(0.107)	
δ			0.908*
			(0.052)
β_1			0.581***
			(0.067)
β_0			0.770***
			(0.066)
P-Value ($\delta = 1$)	0.2598	0.1444	0.0790
P-Value ($\beta_1 = 1$)	0.0000	0.0001	0.0000
P-Value ($\beta_0 = 1$)	0.0053	0.0255	0.0006
Diff ($\beta_1 - \beta_0$)	-0.191	-0.179	-0.188
P-Value ($\beta_1 = \beta_0$)	0.0219	0.2074	0.0118
R^2	0.59	0.65	0.61
Observations	297	183	480
P-Value [Chow-test] for $\delta^V = \delta^N$			0.8400
P-Value [Chow-test] for $\beta_1^V = \beta_1^N$			0.9367
P-Value [Chow-test] for $\beta_0^V = \beta_0^N$			0.8578
P-Value [Chow-test] for $(\beta_1^V - \beta_0^V) - (\beta_1^N - \beta_0^N)$			0.9403

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

Table E2: Updating Beliefs Within Events: 40% ≤ Prior ≤ 60%

Dependent Variable: Logit Posterior Belief					
Regressor	(1) Easy Dice	(2) Hard Dice	(3) Weather	(4) Quiz (S)	(5) Quiz (O)
δ	0.828**	0.832**	1.036	0.768	1.069
	(0.077)	(0.076)	(0.120)	(0.181)	(0.093)
β_1	0.311***	0.337***	0.740**	0.639***	0.552**
	(0.156)	(0.093)	(0.111)	(0.122)	(0.217)
β_0	0.934	0.886	0.782**	0.694***	0.505***
	(0.260)	(0.178)	(0.088)	(0.110)	(0.126)
P-Value ($\delta = 1$)	0.0366	0.0344	0.7676	0.2093	0.4682
P-Value ($\beta_1 = 1$)	0.0002	0.0000	0.0216	0.0052	0.0596
P-Value ($\beta_0 = 1$)	0.8019	0.5254	0.0159	0.0084	0.0018
Diff ($\beta_1 - \beta_0$)	−0.623	−0.549	−0.042	−0.055	0.047
P-Value ($\beta_1 = \beta_0$)	0.0674	0.0032	0.6916	0.6816	0.8545
R^2	0.67	0.62	0.61	0.47	0.80
Observations	64	81	189	111	35

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

Table E3: Priors Greater or Less than One Half

Dependent Variable: Logit Posterior Belief			
Regressor	(1) Prior > $\frac{1}{2}$	(2) Prior < $\frac{1}{2}$	(3) All
δ	0.899***	0.987	0.914***
	(0.016)	(0.036)	(0.009)
β_1	0.542***	0.466***	0.588***
	(0.049)	(0.064)	(0.034)
β_0	0.819***	0.888	0.793***
	(0.058)	(0.069)	(0.038)
P-Value ($\delta = 1$)	0.0000	0.7085	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_0 = 1$)	0.0021	0.1072	0.0000
Diff ($\beta_1 - \beta_0$)	−0.277	−0.422	−0.205
P-Value ($\beta_1 = \beta_0$)	0.0011	0.0001	0.0000
R^2	0.69	0.65	0.84
Observations	2253	927	3360

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant. First column includes updates in direction predicted by Bayes' rule. Second column replaces boundary probabilities with 0.01 or 0.99 respectively. Third column is entire sample.

E.2 Additional Results for Investigating Signal Structure

Table E5 presents the analogous analysis to Table 3 in the paper, but omitting the Quiz (Self) event, since signals regarding the quiz event depend on ability, a potential confound. Regarding Columns 1 to 4, the results are by and large unchanged. Regarding Columns 5 and 6, if anything, the results present even stronger evidence of differential asymmetry between those who received exactly the same sequence of signals (1 affirmative and 1 negative), but only differed in the order these were received. In Column 5 the negative asymmetry is significant at the 5% level, while in Column 6 the positive asymmetry is not significant at conventional levels. However, the difference in the asymmetry is statistically significant at the 5% level (Chow Test).

The result that levels of asymmetry in Columns 5 and 6 are different, solely based on the *order* of signals received is highly surprising. This result is not driven by differences in the average or even the distribution of prior beliefs for these individuals. Table E4 presents tests of equality for the prior beliefs used in Tables 3 and E5 (note these are updated beliefs after receiving two rounds of signals). From these tests, one can see that prior beliefs are quite similar, as one would expect given individuals who received identical signals in the past. Excluding the quiz, in fact leads to slightly improved balance across the two groups.

Table E4: Comparing Beliefs for Individuals in Columns 5 and 6 in Tables 3 and E5

Equality tests of prior beliefs		1st '−'; 2nd '+'	1st '+'; 2nd '−'	Difference
Incl. Quiz (Table 3)	Mean	0.346	.3733	-0.027
	Median	0.250	0.260	-0.010
	Std. Dev.	0.274	0.300	
	Observations	289	270	
	Wilcoxon rank-sum (p-value)			0.450
	Kolmogorov-Smirnov (p-value)			0.365
Excl. Quiz (Table E5)	Mean	0.323	.334	-0.011
	Median	0.250	0.200	-0.050
	Std. Dev.	0.266	0.290	
	Observations	243	215	
	Wilcoxon rank-sum (p-value)			0.880
	Kolmogorov-Smirnov (p-value)			0.841

Table E6 presents additional specifications intended to examine the observed bias related to signal structure discussed in Section 4.4. Columns 1 and 2 present updating in the second round, after individuals had received two signals in total. It separates those who received a negative (−) signal as their previous (first) signal (Column 1), with those who received previously an affirmative (+) signal (Column 2). Columns 3 and 4 present analogous regressions for updating in the third and final round, given the previous (second) signal.

Confirming earlier observed patterns, and contrary to the Bayesian prediction, Table E6 shows a significant negative asymmetry in Column 1 (following a negative signal), and a positive (though not significant) asymmetry in Column 2 (following an affirmative signal). In Columns 3 and 4 the asymmetry is negative in both cases, though it is worth noting that the difference in asymmetry between the two regressions is of similar magnitude.[11]

Table E5: Updating Beliefs in Final Round By Distribution of Signals Received (Excluding Quiz (self))

Dependent Variable: Logit Posterior Belief						
Regressor	(1) 0 '+' Signals	(2) 1 '+' Signal	(3) 2 '+' Signals	(4) 3 '+' Signals	(5) 1st '−'; 2nd '+'	(6) 1st '+'; 2nd '−'
δ	0.899***	0.890***	0.915**	0.988	0.860***	0.906***
	(0.033)	(0.025)	(0.038)	(0.077)	(0.035)	(0.036)
β_1		0.323***	0.903	1.244	0.750**	0.973
		(0.103)	(0.090)	(0.171)	(0.117)	(0.158)
β_0	1.106	0.920	0.659***		1.090	0.757**
	(0.118)	(0.081)	(0.124)		(0.113)	(0.102)
P-Value ($\delta = 1$)	0.0023	0.0000	0.0261	0.8820	0.0001	0.0097
P-Value ($\beta_1 = 1$)		0.0000	0.2817	0.1588	0.0339	0.8644
P-Value ($\beta_0 = 1$)	0.3719	0.3273	0.0064		0.4264	0.0183
Diff ($\beta_1 - \beta_0$)		−0.597	0.244		−0.340	0.216
P-Value ($\beta_1 = \beta_0$)		0.0001	0.1053		0.0439	0.2502
R^2	0.77	0.81	0.76	0.76	0.82	0.81
Observations	206	380	220	60	225	199

Analysis uses OLS regression. Columns (1)-(4): K '+ Signals' refers to K affirmative signals, out of a possible maximum of 3. Columns (5)-(6): Compares individuals who received exactly 1 affirmative and 1 negative signal, only differing in the order these signals were received. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

[11] Given the patterns observed here and in Section 4.4, more negative asymmetry is to be expected in the third rather than second round, as the average proportion of negative signals received is greater.

Table E6: Updating Beliefs by Sequences of Signals Received

	Dependent Variable: Logit Posterior Belief			
	After 2nd Signal (Round 2)		After 3rd Signal (Round 3)	
Regressor	(1) 1st Signal '−'	(2) 1st Signal '+'	(3) 2nd Signal '−'	(4) 2nd Signal '+'
δ	0.910***	0.913***	0.890***	0.878***
	(0.019)	(0.023)	(0.026)	(0.025)
β_1	0.372***	0.746***	0.328***	0.811**
	(0.053)	(0.082)	(0.091)	(0.083)
β_0	0.887*	0.679***	1.142	0.971
	(0.060)	(0.071)	(0.094)	(0.079)
P-Value ($\delta = 1$)	0.0000	0.0002	0.0000	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0022	0.0000	0.0234
P-Value ($\beta_0 = 1$)	0.0590	0.0000	0.1332	0.7139
Diff ($\beta_1 - \beta_0$)	−0.514	0.067	−0.814	−0.160
P-Value ($\beta_1 = \beta_0$)	0.0000	0.5398	0.0000	0.1499
R^2	0.83	0.83	0.78	0.81
Observations	700	435	377	515

Analysis uses OLS regression. '+' refers to affirmative signal, '−' to negative. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

E.3 Excluding the Last Round

Table E7 presents the analogous analysis to Table 1 in the main paper, excluding the last round. The motivation for excluding the last round of updating is to understand how much of the asymmetry in the data could be explained by representativeness bias. In the final round, subjects have potentially observed one of the two sets of "representative" signal sequences - i.e. sequences that exactly match the strength of signals. Two negative and one affirmative, exactly matches the expected number of signals for an event that did not occur, while two affirmative and one negative exactly matches the expected number of signals for an event that did occur. The representativeness bias would be to bias the posterior towards 0 in the first case, and towards 1 in the second case. In the framework this could be manifested as an exaggerated response to signals that go in the majority direction, and a conservative response to signals that go against the majority. By eliminating the final round, subjects cannot make use of the representativeness heuristic.

From Table E7 it is possible to see that the observed negative asymmetry persists in earlier updating rounds. Thus, while representativeness bias may play a role, it cannot explain the negative asymmetry observed in the data.

Table E7: Updating Beliefs for All Events: Rounds 1-2 only

Dependent Variable: Logit Posterior Belief			
Regressor	(1) Good/Bad News	(2) Just News	(3) All
δ^V	0.910***		
	(0.014)		
β_1^V	0.528***		
	(0.042)		
β_0^V	0.722***		
	(0.045)		
δ^N		0.895***	
		(0.017)	
β_1^N		0.454***	
		(0.050)	
β_0^N		0.762***	
		(0.060)	
δ			0.904***
			(0.011)
β_1			0.500***
			(0.035)
β_0			0.736***
			(0.041)
P-Value ($\delta = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_0 = 1$)	0.0000	0.0001	0.0000
Diff ($\beta_1 - \beta_0$)	−0.194	−0.308	−0.236
P-Value ($\beta_1 = \beta_0$)	0.0002	0.0000	0.0000
R^2	0.85	0.83	0.84
Observations	1343	972	2315
P-Value [Chow-test] for $\delta^V = \delta^N$			0.4444
P-Value [Chow-test] for $\beta_1^V = \beta_1^N$			0.2195
P-Value [Chow-test] for $\beta_0^V = \beta_0^N$			0.5313
P-Value [Chow-test] for $(\beta_1^V - \beta_0^V) - (\beta_1^N - \beta_0^N)$			0.1709

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

F Sampling Robustness Checks

F.1 Sample Restrictions

Table F1 examines the impact of how restricting the sample alters updating estimates in the main framework. The first column presents the main analysis (Column 3 in Table 1), but includes observations where belief updates go in the opposite direction that Bayes' rule predicts. The second column replaces boundary observations of 0 or 1 with 0.01 or 0.99 respectively. In Table 1 these were dropped. Finally the third column also truncates boundary observations, and includes updates in the wrong direction. The third column thus presents the full data, with no exclusions.

Table F1: Relaxing Sample Restrictions and Full Sample

Dependent Variable: Logit Posterior Belief			
Regressor	(1) Include Wrong Dir.	(2) Include Boundary	(3) Include All
δ	0.910***	0.914***	0.914***
	(0.010)	(0.010)	(0.011)
β_1	0.506***	0.727***	0.649***
	(0.033)	(0.045)	(0.045)
β_0	0.714***	0.903**	0.805***
	(0.038)	(0.045)	(0.045)
P-Value ($\delta = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_1 = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_0 = 1$)	0.0000	0.0306	0.0000
Diff ($\beta_1 - \beta_0$)	−0.208	−0.176	−0.156
P-Value ($\beta_1 = \beta_0$)	0.0000	0.0003	0.0022
R^2	0.81	0.81	0.79
Observations	3537	3654	3840

Analysis uses OLS regression. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant. First column includes updates in direction predicted by Bayes' rule. Second column replaces boundary probabilities with 0.01 or 0.99 respectively. Third column is entire sample.

F.2 Restricting to Active Updates

Table F2 presents the analysis of Table 1, but restricting the sample to only active updates. The results show that subjects appear to suffer from the opposite bias of conservatism, as they are over-responsive to information. This is largely drive by response to a negative signal, but does not appear to differ between good or bad news, versus just news. As such, symmetry can be rejected at the 1% level.

Table F2: Active Updates: Reponse to Contemporaneous Signal

Dependent Variable: Logit Posterior Belief			
Regressor	(1) Good/Bad News	(2) Just News	(3) All
δ	0.882***	0.863***	0.873***
	(0.018)	(0.022)	(0.014)
β_1	1.060	1.092	1.074
	(0.052)	(0.071)	(0.047)
β_0	1.295***	1.323***	1.305***
	(0.050)	(0.071)	(0.046)
P-Value ($\delta = 1$)	0.0000	0.0000	0.0000
P-Value ($\beta_1 = 1$)	0.2529	0.1962	0.1138
P-Value ($\beta_0 = 1$)	0.0000	0.0000	0.0000
Diff ($\beta_1 - \beta_0$)	−0.235	−0.232	−0.231
P-Value ($\beta_1 = \beta_0$)	0.0002	0.0121	0.0000
R^2	0.81	0.79	0.80
Observations	1121	799	1920

Analysis uses OLS regression. Includes only active updates. Difference is *significant from 1* at * 0.1; ** 0.05; *** 0.01. Robust standard errors clustered at individual level. R^2 corrected for no-constant.

G Aggregate Updating by Event/Stake/Accuracy Payment

In this section I examine patterns in updating behavior for different events and financial stake conditions. Recall that the lump sum payment used for the lottery method was randomized at the session level, and was either $3, $10, or, $20. The financial stake was randomized at the individual-event level, and was either $0 or $80 with 50% probability respectively. The financial stake was an amount of money that would be gifted to the subject if the event occurred and had been randomly selected for payment.

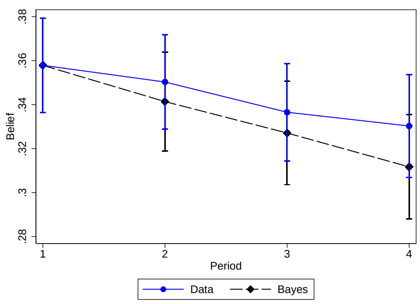
In Figure G1 I examine the analog to Figure 4, for each of the two financial stake conditions ($0 and $80), as well as each of the three accuracy payment conditions ($3, $10, $20). While different values of the accuracy payment do not affect whether news is good or bad, note that having an $80 stake in an event necessitates that signals contain either good or bad news.

From Figure G1 there does not appear to be any sizeable differences in updating behavior across these different payment conditions. The results on differences between a stake of $0 versus $80 are consistent with Barron (2016), who does not find evidence of asymmetry when individuals have a financial stake in an event. Note also that the prior varies slightly by payment conditions; updating patterns by prior are presented in Figure E1 below.

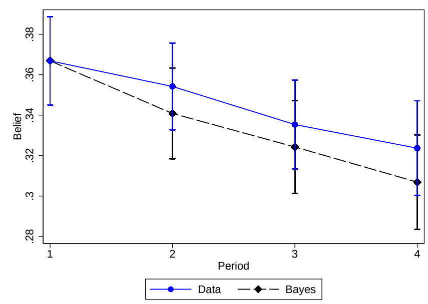
Next, in Figure G2 I present the analogous analysis for each of the four events, with the quiz event split into the self and other treatments. For the two dice events, which involved the probability that particular outcomes from rolls of either two or four dice had occurred, updating appears to be more conservative than the aggregate. The pattern is also seen when individuals estimate the probability that another randomly selected, anonymous individual in the room had scored in the top 15% on the earlier taken quiz (quiz: other performance).

For the quiz (self performance) event, which involved the probability that the individual believed they scored in the top 15% of quiz takers, updating appears to adhere more closely to the Bayesian prediction. This is also true for the weather event, which occurred when subjects had correctly estimated the mean temperature $\pm$ 5 degrees F in New York City on a randomly selected day in the previous calendar year. In the aggregate, updating about own performance does not appear to deviate much from the Bayesian prediction.

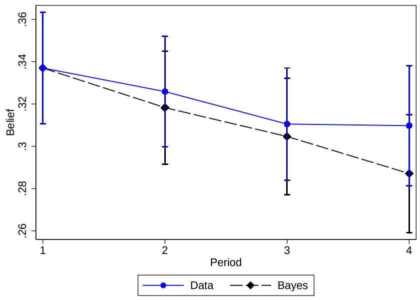
Figure G1: Evolution of Beliefs By Stake and Accuracy Conditions



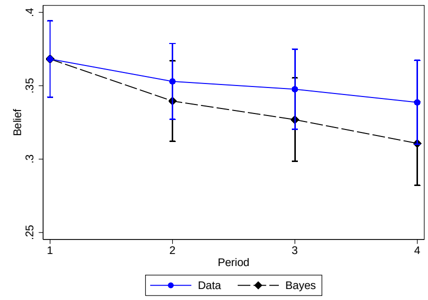
(a) Stake = \$0



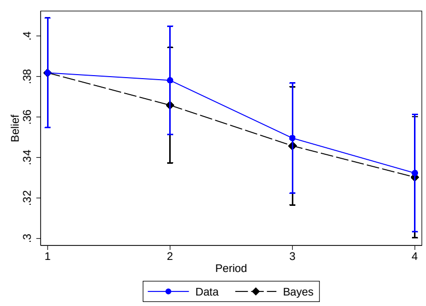
(b) Stake = \$80



(c) Accuracy Payment = \$3



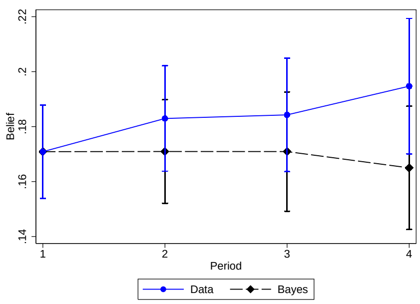
(d) Accuracy Payment = \$10



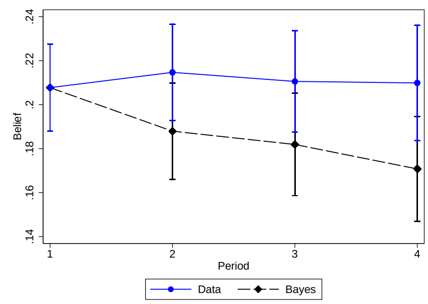
(e) Accuracy Payment = \$20

The path of beliefs starting from the prior (period 1), and after each sequential signal (periods 2 through 4). Average individual responses are the blue solid line, the Bayesian benchmark is marked as the black dashed line. Bayesian benchmark takes prior beliefs, and subsequently uses Bayes' rule to update beliefs. Error bands represent 95% confidence intervals. Note the potential difference in the range of prior beliefs, on the vertical axis. $N = \{646, 634, 424, 436, 420\}$ per round, respectively for (a)-(e).

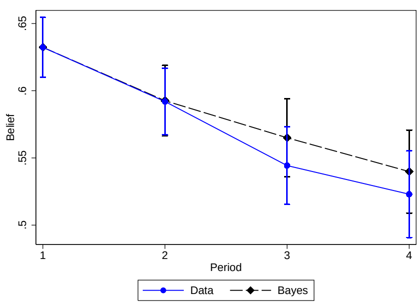
Figure G2: Evolution of Beliefs: By Event



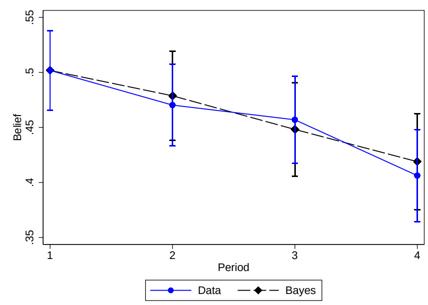
(a) Easy Dice



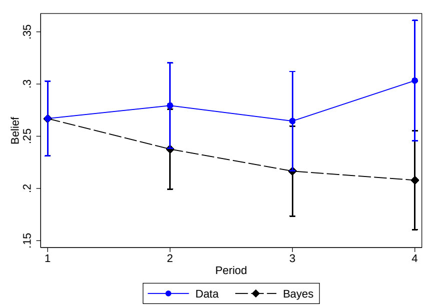
(b) Hard Dice



(c) Weather Event



(d) Quiz Event (self performance)



(e) Quiz Event (other performance)

The path of beliefs starting from the prior (period 1), and after each sequential signal (periods 2 through 4). Average individual responses are the blue solid line, the Bayesian benchmark is marked as the black dashed line. Bayesian benchmark takes prior beliefs, and subsequently uses Bayes' rule to update beliefs. Error bands represent 95% confidence intervals. Note the difference in the range of prior beliefs, on the vertical axis. $N = \{318, 318, 326, 223, 95\}$ per round, respectively for (a)-(e).

H Experiment Instructions

Note: Accuracy payments were randomized at the session level $a \in \$\{3, 10, 20\}$*. Instructions show $20 for exposition only.*

Instructions (Section 1)

Thank you for your participation in this experiment! This experiment will last approximately 80 minutes. This experiment is about how likely you think an uncertain event is to have occurred. You will consider four such separate events today, which will be presented one at a time. For these events, we want you to think in terms of the percent chance out of 100 that they occurred. For example, you may believe that there is 50% chance that when flipping a coin it will come up TAILS. This experiment has been designed so that you have the greatest chance of earning the most money when you carefully and accurately think about the percent chance of such an event occurring.

You will be awarded a $10 show-up fee for your participation until the end, in addition to anything you may earn during the experiment. Please also note the following during the experiment:

- Please put away any cell phones/devices. Outside communication or accessing the internet during this experiment is forbidden. Violators will not receive payment and will be blacklisted from the lab.
- Please do not communicate with others in the lab, except to ask questions
- If you have a question please do not hesitate to ask! Questions are encouraged!

We will now introduce the experiment through Instructions 1-3 and three short practice sessions that go with each set of instructions. The practice sessions are to help you get familiar with the experiment's components that will ALL be combined when doing the final experiment for money.

The "Main Event"

In this experiment you are estimating the percent chance that a "main event" occurred. An example of a "main event" is: the average temperature in the contiguous USA was warmer in 2013 than 2012. Your earnings are in part based on the accuracy of your predictions of whether the "main event" occurred. Think about the following: What is the probability the average temperature in the USA was warmer in 2013 than 2012?

How will I record my percent chance estimate?

First we introduce a gumball machine with 100 green and black gumballs. For example, suppose there are 40 green and 60 black gumballs. Most people would agree that the probability of drawing a green gumball is exactly 40%. Now think back to the "main event" about the weather being warmer in 2013 than 2012 in the US. We next give you $20. But this $20 must be wagered on one of two scenarios.

1. The “gumball event”: Drawing a green gumball from a machine with 40 out of 100 green, OR

2. The “main event”: the average US temperature in 2013 was warmer than it was in 2012.

You have to decide if you think the chance that the weather was warmer in 2013 is greater than 40%, or less than 40%. If you decide to wager the $20 on the “gumball event”, the computer randomly draws a gumball from the machine with 40 green (60 black) gumballs. If it’s green you win the $20. If black, you get nothing. If you decided to go with the “main event”: the climate being warmer in 2013, we check the statistics. If it was warmer, you win the $20. If it was colder, you get nothing.

Consider different numbers of green gumballs:

If the gumball machine has only 2 green gumballs (98 black) would you prefer to wager $20 on the “gumball event” or the “main event”? Most of you probably think the climate being warmer in 2013 than 2012 is more likely than 2% and prefer to wager the $20 on the “main event”.

What if the gumball machine has 25 green gumballs? Those who think the “main event” is more likely than 25% would want to wager on the “main event”. Now, what if the gumball machine has 90 green gumballs? The “gumball event” now pays off with 90% chance. Probably, almost everyone will prefer to wager the $20 on the gumball machine, except for those that think there is a greater than 90% chance that the weather was warmer in 2013.

Example – You think there is a 35% chance the weather is warmer in 2013 than 2012.

- Case 1: Whenever you see a gumball machine with 34 or less green gumballs, to earn the most money you would want to wager the $20 on the “main event”. E.g. if there were 5 green gumballs, 5% is a lower chance than 35% of earning the $20.

- Case 2: If you see a gumball machine with 36 or more green gumballs, you would prefer to wager the $20 on the “gumball event”. E.g. If there were 60 green gumballs, this is a 60% chance of drawing green – better than the 35% chance you think the weather would be warmer.

- If there are exactly 35 green gumballs, you probably don’t care whether to wager your $20 on the “gumball event” or the “main event”. Both give you a 35% chance of earning the $20.

The “Slider”

In this experiment you are going to indicate on a “slider” exactly how many gumballs need to be green before you prefer to wager $20 on the “gumball event” instead of some other “main event”. In other words, you will indicate the minimum number of gumballs that have to be green, before you prefer to wager $20 on the gumball machine. To make sure it is in your best financial interest to do this, after you have made your slider choice we are going to randomly fill a gumball machine with 0 to 100 green gumballs and the rest black. Each possible number of green gumballs is equally likely – and your slider choice has no effect on the number chosen. Based on your slider choice, we will then make the $20 wager for you. If there happen to be less green gumballs than the minimum you chose, your $20 is wagered on whatever main event you are predicting. If there happen to be more (or the same) green gumballs than the minimum you indicated in the slider, we will wager your $20 on drawing a green gumball from this machine we randomly filled.

If this is a little confusing, you can just remember, to have the highest chance of earning money, your slider choice should be exactly the probability out of 100 you think the event has of occurring.

Summary of Section 1

- Make selection on the "Slider" for your estimate of the "main event"
- Computer randomly generates an amount (out of 100) of "green gumballs"
- The amount of green gumballs determines how the $20 is wagered in your best interest. 1) The "main event" or 2) The "gumball event". The outcome of the $20 wager is then revealed.

Are there any questions?

Instructions (Section 2) – "Feedback"

Now we're going to make things more interesting. Suppose now the "Main event" is that the average temperature in 1998 was warmer than 1997 in the contiguous USA.

Please note – these events are used for practice. The real events may (and will) be different.

You will again adjust the slider to indicate how likely you believe this is to be true. But now, after you adjust the "Slider" the first time, you are going to get some "feedback" about whether or not 1998 was in fact warmer than 1997.

What is "Feedback"?

"Feedback" is information about the main event that gives you additional clues to help you make your selection. Please note that you are provided three rounds of this "feedback" – however each time you are presented with this "feedback" it may or may not be telling you the truth. For our experiment we use gremlins to provide the three rounds of feedback when making your selection. For each round, two gremlins always tell the truth while one of them, Larry, always lies. You will not know which gremlin is talking and after you get this "feedback", you can adjust your prediction on the 'Slider" if you choose to use their information. Note: The gremlins are randomly chosen "with replacement", meaning that every time you get "feedback" it is true with 2/3 probability. This means, that it's even possible (though unlikely) that all three rounds of feedback come from the gremlin that lied!

Remember: All 3 gremlins always know whether the event happened or not. It's just that only 2 of these 3 tell the truth. When we determine your earnings, before filling the gumball machine we are going to randomly only pick one of these four slider choices. Are there any questions at this point? Next we proceed to the second practice. In this example please note two additional tools for your use.

1. Calculate Fraction: Pulls up a calculator in case you want to transform a fraction to a decimal.

2. Show History: Shows you your history of feedback from gremlins AND your past slider choices.

Instructions (Section 3) – Payment groups

The last component explains how you might earn additional money during this experiment. This is very important to understand when conducting the final experiment. You will all be in one of two payment groups: "red" or "blue". NOTE: You will not know which payment group (red or blue) you are in when you make your slider choices. Suppose now the 'main event" is whether the climate in the USA was warmer in 1990 than 1980.

"The Red Group"

Half of you are going to be in the "red" group. In the "red" group, your payment at the end looks exactly like how we have been practicing so far. We will pick one of your four slider choices incorporating the "feedback", and then fill a gumball machine with a random number of green gumballs. Based on your selection, if the $20 is wagered on the "gumball event" then a gumball would be drawn – if green you earn the $20. If the $20 is wagered on the "main event", then if that event occurred you earn the $20.

"The Blue Group"

The other half of you will be in the "blue" group. The "blue" group automatically gets $20, just for being blue. In this group, the slider choices previously selected do not matter for payment. Instead payment depends on a "blue bonus chip" provided that pays out only if the event you are predicting actually occurs. Taking the example of climate, if 1990 was warmer than 1980, and if you are in the blue group, you would receive $20 automatically, plus whatever amount is on the "blue bonus chip". The amount on the chip is either $0 or $80. Each is equally likely. Example: If you're in the "blue" group you would automatically earn $20, and if the main event you are predicting occurs you would also earn the amount on the blue bonus chip ($0 or $80): for a maximum earnings of $100.

"Blue Bonus chip"

Everyone will get a "blue bonus chip" prior to knowing which group you are in and prior to each of the four events. The experiment coordinator will fill a bag with half $0 chips and half $80 chips. Then, each of you will draw one of these chips from the bag. Note that having a "blue bonus chip" is only significant when you end up in the "blue" group and indicates how much is earned if the event happens AND if you are in the "blue" group.

Each of you has a fair, 50% chance of drawing an $80 bonus chip. There is no advantage to drawing a chip earlier or later, everyone in this room has the same 50% chance. Even if you are the last to draw, and there is only one chip left, that one chip is $0 with 50% chance and $80 with 50% chance. Since you don't know if you're "red" or "blue" until all slider choices have been made, in order to have the best chance of earning the most money, it pays to be as accurate as possible when making slider choices.

Are there any questions at this point? Next we proceed to the final practice. Note that your "blue bonus chip" has an 8-digit code that you are required to enter into the computer. Your "blue bonus chip" does not affect in any way the event that you will be predicting. The event is the same if you pick a $0 chip or an $80 chip. Forget about the gremlins or "feedback" for this practice, yet they will be in the main experiment.

Summary for the Final experiment

Now we are ready to put ALL the pieces together for the final experiment! There are going to be four main events, however only one will be picked at random for payment.

1. The coordinator will come around with a bag that contains a 50/50 mix of $0 and $80 "blue bonus chips" for the upcoming event.

2. Make a note of your "blue bonus chip" amount. This is what you could earn if the event happens AND if you also happen to be in the blue group.

3. The event will be described to you. Next, indicate on the "Slider" the probability you believe the event occurred. Your slider choice does not affect how many green gumballs the random gumball machine will have nor does it affect the chances of the "main event".

4. You'll get "Feedback" three times from a random gremlin. Remember there is a 2/3 chance the feedback is true. You can choose to use this information if you want to reassess the probability by indicating this on the slider after each "Feedback".

5. Steps 1 to 4 are repeated for each of the four events.

After making all of your slider choices:

1. The coordinator will come with two bags. The color bag contains 50/50 mix of blue and red chips. The chip you draw determines if your payment group is red or blue. If it is red, the slider choice (1-4) is indicated on the chip.
2. The event bag contains an equal amount of Event #1, #2, #3 and #4 chips. The number on the chip determines what event will be paid.

Suppose you picked the chip for Event #1.

1. IF draw RED: The chip indicates the slider choice. A gumball machine is filled with a random number of green gumballs. Based on your slider choice, \$20 is wagered on gumball machine or Event #1, as we practiced.
2. IF draw BLUE: The outcome of Event #1 is revealed. If the event occurred you earn \$20 + the amount on your event #1 bonus chip, \$80 or \$0. If the event did not occur you just earn the \$20. After your payment is determined, we will reveal the outcomes of the other three events. This is for your information only, and it does not affect your payment.

Important Notes:

The procedures that will occur today have been approved by the University Committee on Activities Involving Human Subjects (UCAIHS). This experiment complies with UCAIHS requirements (HS# 10-8117), in particular, not to engage in any deception or misinformation about the probabilities presented today.

- When you encounter random chance off the computer (e.g. when drawing chips from the bag) we make every effort to ensure that this is transparent and legitimate. If we state there is a 50-50 chance of drawing a particular chip, we will have at least one participant verify that this is indeed the case. (any participant may ask to verify the bag contents before the draws begin)
- When you encounter random chance on the computer (e.g. drawing a gumball from a hypothetical machine) the computer has been programmed to perform the randomization exactly as is stated in this experiment. For example, if you are told that there are 30 green gumballs and 70 black, the computer is programmed to randomly select a green gumball with exactly 30 chances out of 100.

Before moving forward to the next main event, the computer will wait for everyone to finish the current event. There is no advantage to finishing quickly, as you will end up waiting for other participants.