Supplementary Materials for:

Different Methods Elicit Different Belief Distributions

Beidi Hu[1] and Joseph P. Simmons[2]

[1] The University of Chicago Booth School of Business

[2] The Wharton School, University of Pennsylvania

Table of Contents

Supplement 1: Exclusions and Reported Sample Sizes

Table S1 provides the full breakdown of the target sample sizes, pre-registered exclusions, and reported sample sizes for Studies 1-10, S1-S3.

Table S1. Exclusions in Studies 1-S3

Study 1: Starting N = 1,479

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	84	1,395		
failed the comprehension check twice	27	1,368		
dropped out of the survey before being assigned to a condition	13	1,355		
dropped out of the survey after being assigned to a condition	53	1,302	34	19
were associated with a duplicate IP Address	22	1,280	11	11
were associated with a duplicate worker ID	2	1,278	1	1
provided response with an average absolute error of 0	2	1,276	2	0

Final sample: N = 1,276: 630 in the Distribution Builder condition; 646 in the Sliders condition

Study 2: Starting N = 1,115

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder first condition	# excluded from Sliders first condition
dropped out of the survey before comprehension check	102	1,013		
failed the comprehension check twice	12	1,001		
dropped out of the survey after comprehension check	0	1,001	0	0
were associated with a duplicate IP Address	19	982	6	13
were associated with a duplicate worker ID	7	975	3	4
	Total # of excluded observations	# of remaining observations	# of observations excluded from Distribution Builder condition	# of observations excluded from Sliders condition
provided response with an average absolute error of 0	8	1,942	6	2

Final sample: 1,942 observations (coming from 974 participants):
969 observations for Distribution Builder; 973 observations for Sliders

Note. We manipulated method within-subject in this study, such that participants used Distribution Builder in one round and Sliders in another round. And we counterbalanced the order.

Study 3: Starting N = 1,647

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder first condition	# excluded from Sliders first condition
dropped out of the survey before comprehension check	29	1,618		
failed the comprehension check twice	9	1,609		
dropped out of the survey after comprehension check	0	1,609	0	0
were associated with a duplicate IP Address	9	1,600	6	3
were associated with a duplicate worker ID	2	1,598	2	0
misreported worker ID	8	1,590	3	5
	Total # of excluded observations	# of remaining observations	# of observations excluded from Distribution Builder condition	# of observations excluded from Sliders condition
provided response with an average absolute error of 0	16	3,164	7	9

Final sample: 3,164 observations (coming from 1,588 participants):

1,583 observations for Distribution Builder; 1,581 observations for Sliders

Note. We manipulated method within-subject in this study, such that participants used Distribution Builder in one round and Sliders in another round. And we counterbalanced the order.

Study 4: Starting N = 2,088

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	96	1,992		
failed the comprehension check twice	117	1,875	60	57
dropped out of the survey after comprehension check	173	1,702	98	75
were associated with a duplicate IP Address	122	1,580	58	64
were associated with a duplicate worker ID	3	1,577	1	2
	Total # of excluded observations	# of remaining observations	# of observations excluded from Distribution Builder condition	# of observations excluded from Sliders condition
provided response with an average absolute error of 0	31	3,123	19	12

Final sample: 3,123 observations (coming from 1,573 participants):
1541 observations in the Distribution Builder condition; 1,582 observations in the Sliders condition

Study 5: Starting N = 1,773

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before being assigned to a condition	0	1,773		
dropped out of the survey after being assigned to a condition	65	1,708	32	33
were associated with a duplicate IP Address	5	1,703	2	3
were associated with a duplicate worker ID	0	1,703	0	0
failed any attention check in the survey	167	1,536	78	89
provided the same answers to all questions in part 1 of the survey	0	1,536	0	0

Final sample: N = 1,536: 774 in the Distribution Builder condition; 762 in the Sliders condition

Study 6: Starting N = 2,340

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	267	2,073		
failed the comprehension check twice	26	2,047	13	13
dropped out of the survey after comprehension check	48	1,999	24	24
were associated with a duplicate IP Address	91	1,908	34	57
were associated with a duplicate worker ID	2	1,906	2	0
	Total # of excluded observations	# of remaining observations	# of observations excluded from Distribution Builder condition	# of observations excluded from Sliders condition
provided response with an average absolute error of 0	11	3,801	8	3

Final sample: 3,801 observations (coming from 1,905 participants):
1,920 observations in the Distribution Builder condition; 1,881 observations in the Sliders condition

Study 7: Starting N = 2,179

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	302	1,877		
failed the comprehension check twice	168	1,709	87	81
dropped out of the survey after comprehension check	49	1,660	30	19
were associated with a duplicate IP Address	93	1,567	46	47
were associated with a duplicate worker ID	3	1,564	1	2

Final sample: N = 1,564: 761 in the Distribution Builder condition; 803 in the Sliders condition

Study 8: Starting N = 1,292

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before attention check	85	1,207		
failed the attention check	8	1,199	7	1
dropped out of the survey after attention check	0	1,199	0	0
were associated with a duplicate IP Address	15	1,184	8	7
were associated with a duplicate worker ID	4	1,180	2	2
misreported worker ID	18	1,162	8	10

Final sample: N = 1,162: 582 in the Distribution Builder condition; 580 in the Sliders condition

Study 9: Starting N = 1,214

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before attention check	3	1,211		
failed the attention check	10	1,201	6	4
dropped out of the survey after attention check	0	1,201	0	0
were associated with a duplicate IP Address	12	1,189	5	7
were associated with a duplicate worker ID	0	1,189	0	0
misreported worker ID	4	1,185	2	2

Final sample: N = 1,185: 594 in the Distribution Builder condition; 591 in the Sliders condition

Study 10: Starting N = 2,608

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Regular Sliders condition	# excluded from Reverse Sliders condition
dropped out of the survey before comprehension check	345	2,263			
failed the comprehension check twice	248	2,015			
dropped out of the survey after comprehension check	36	1,979	18	13	5
were associated with a duplicate IP Address	182	1,797	59	52	71
were associated with a duplicate worker ID	4	1,793	2	2	0
provided response with an average absolute error of 0	3	1,790	0	2	1

Final sample: N = 1,790: 592 in the Distribution Builder condition; 602 in the Regular Sliders condition; 593 in the Reverse Sliders condition

Study S1: Starting N = 3,375

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Horizontal Sliders condition	# excluded from Vertical Sliders condition
dropped out of the survey before attention check	361	3,014			
failed the attention check	20	2,994	10	5	5
dropped out of the survey after attention check	0	2,994			
were associated with a duplicate IP Address	59	2,935	18	22	19
were associated with a duplicate worker ID	11	2,924	3	2	6
misreported worker ID	33	2,891	12	9	12

Final sample: N = 2,891: 960 in the Distribution Builder condition; 975 in the Horizontal Sliders condition; 956 in the Vertical Sliders condition

Study S2: Starting N = 1,078

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	76	1,002		
failed the comprehension check twice	2	1,000	2	0
dropped out of the survey after comprehension check	0	1,000	0	0
were associated with a duplicate IP Address	18	982	3	15
were associated with a duplicate worker ID	0	982	0	0
misreported worker ID	3	979	2	1
provided response with an average absolute error of 0	3	976	3	0

Final sample: N = 976: 490 in the Distribution Builder condition; 486 in the Sliders condition

Study S3: Starting N = 1,260

Reason for exclusions	Total # of exclusions	# of remaining participants	# excluded from Distribution Builder condition	# excluded from Sliders condition
dropped out of the survey before comprehension check	30	1,230		
failed the comprehension check twice	28	1,202	14	14
dropped out of the survey after comprehension check	0	1,202	0	0
were associated with a duplicate IP Address	13	1,189	7	6
were associated with a duplicate worker ID	4	1,185	2	2
provided response with an average absolute error of 0	2	1,183	2	0

Final sample: N = 1,183: 587 in the Distribution Builder condition; 596 in the Sliders condition

Table S1.2 provides detailed demographic information for Studies 1-7 individually.

Table S1.2 Demographics Information for Studies 1-7

	Final sample (N)	Reported "Female"	Reported "Male"	Age
Study 1	1,276	565 (44.3%)	711 (55.7%)	M = 38.0, SD = 12.3
Study 2	974	488 (50.1%)	486 (49.9%)	M = 42.4, SD = 13.6
Study 3	1,588	841 (53.0%)	747 (47.0%)	M = 39.6, SD = 13.4
Study 4	1,573	704 (44.8%)	869 (55.2%)	M = 38.6, SD = 12.2
Study 5	1,536	819 (53.3%)	717 (46.7%)	M = 32.8, SD = 11.7
Study 6	1,905	1,019 (53.5%)	886 (46.5%)	M = 32.8, SD = 12.4
Study 7	1,564	798 (51.0%)	766 (49.0%)	M = 38.0, SD = 12.3

Supplement 2: Studies 1-10 Additional Results

In Supplement 2, we report the results of additional accuracy measures (Section 2.1), additional mechanism measures (Section 2.2), moments of elicited distributions (Section 2.3), and comparisons to intuitive benchmarks (Section 2.4).

Section 2.1: Results of Additional Accuracy Measures

In the paper, we discussed the accuracy results with our pre-registered measures of accuracy (Studies 1-6 and 10: average absolute error; Study 7: mass allocated to the correct bin). To ensure the robustness of these findings, we conducted exploratory analyses using additional measures: Kolmogorov-Smirnov distance (KS distance) in Studies 1-6 and 10 and a measure of aggregate deviation in the full distribution in Study 7. Smaller numbers on both measures indicate greater accuracy. The results don't differ meaningfully from our primary measures, and we report them in Table S2.1 below.

In general, we replicated the findings in the paper that Distribution Builder elicited more accurate distributions than Sliders in most cases, and the only exceptions were when the true answer had a right-skewed-looking shape (see the last column in Table S2.1 for an overview).

Table S2.1. Additional Accuracy Measures for Studies 1-7, and 10

	Distribution Builder		Sliders		Effect of the Sliders condition	
Study, Measure, and Distribution	*M*	*SD*	*M*	*SD*	(vs. Distribution Builder condition)	Which method is more accurate?
Study 1						
Kolmogorov-Smirnov Distance[1]						
Right-skewed	0.21	0.12	**0.19**	0.09	b = -0.03, SE = 0.01, p = .005	Sliders
Left-skewed	0.21	0.12	0.22	0.16	b = 0.02, SE = 0.01, p = .280	Distribution Builder
Symmetric	**0.15**	0.10	0.18	0.13	b = 0.03, SE = 0.01, p = .005	Distribution Builder
Study 2						
Kolmogorov-Smirnov Distance						
Right-skewed	0.17	0.09	**0.16**	0.09	b = -0.01, SE = 0.01, p = .029	Sliders
Left-skewed	**0.18**	0.11	0.21	0.13	b = 0.02, SE = 0.01, p < .001	Distribution Builder
Study 3						
Kolmogorov-Smirnov Distance						
Right-skewed	0.16	0.09	0.17	0.09	b = 0.01, SE = 0.004, p = .228	Distribution Builder
Left-skewed	0.18	0.11	0.18	0.12	b = 0.01, SE = 0.01, p = .089	Distribution Builder
Study 4						
Kolmogorov-Smirnov Distance						
Actually Right-skewed / Looks Right-skewed	0.21	0.15	0.20	0.13	b = -0.01, SE = 0.01, p = .330	Sliders
Actually Symmetric / Looks Symmetric	0.21	0.13	0.22	0.15	b = 0.01, SE = 0.01, p = .329	Distribution Builder
Actually Right-skewed / Looks Symmetric	0.19	0.13	0.20	0.12	b = 0.01, SE = 0.01, p = .294	Distribution Builder
Actually Symmetric / Looks Right-skewed	0.21	0.13	0.21	0.14	b = -0.001, SE = 0.01, p = .948	Sliders
Study 5						
Kolmogorov-Smirnov Distance						
Right-skewed	**0.17**	0.09	0.19	0.11	b = 0.02, SE = 0.01, p = .013	Distribution Builder
Symmetric	**0.18**	0.10	0.19	0.13	b = 0.02, SE = 0.01, p = .034	Distribution Builder
Study 6						
Kolmogorov-Smirnov Distance						
Right-skewed	**0.16**	0.09	0.17	0.10	b = 0.01, SE = 0.004, p = .004	Distribution Builder
Left-skewed	**0.17**	0.10	0.19	0.12	b = 0.02, SE = 0.01, p = .001	Distribution Builder
Study 7						
Aggregate Deviation in the Distribution[2]						
Right-skewed	93.89	64.17	95.38	63.60	b = 1.60, SE = 2.38, p = .503	Distribution Builder
Symmetric	**104.20**	55.35	112.55	58.00	b = 8.58, SE = 2.11, p < .001	Distribution Builder
Study 10						
Kolmogorov-Smirnov Distance						
Right-skewed / Regular Sliders	0.20	0.12	0.20	0.12	b = 0.001, SE = 0.01, p = .910	Distribution Builder
Right-skewed / Reverse Sliders	**0.20**	0.12	0.23	0.15	b = 0.03, SE = 0.01, p = .018	Distribution Builder
Left-skewed / Regular Sliders	**0.19**	0.12	0.22	0.14	b = 0.03, SE = 0.01, p = .004	Distribution Builder
Left-skewed / Reverse Sliders	0.19	0.12	0.18	0.11	b = -0.01, SE = 0.01, p = .372	Sliders

Note. Within each row, **boldface** indicates that this method is significantly more accurate than the other method.

[1] Kolmogorov-Smirnov distance is defined as the largest absolute difference between the cumulative distribution function (CDF) of participant's response and the true distribution function. We computed this measure by first deriving the absolute difference between participant's CDF and the true CDF at each response category and then taking the maximum of these absolute differences. Smaller number on this measure indicates higher accuracy.

[2] Because the tasks in Study 7 have a single number rather than a distribution as the correct answer, KS distance cannot be easily computed. To calculate a measure that also takes into account the entire distribution, we first derived the mass participant allocated to each response category, then multiplied each of them by the absolute difference between the category and the category containing the correct answer, and finally summed all these products. We consider the resulting measure as a coarse measure of aggregate deviation across the distribution range. Small number on this measure indicates higher accuracy.

We conducted our pre-registered regression analyses on the additional accuracy measures and obtained functionally the same results. We report them below for interested readers.

Kolmogorov-Smirnov distance (Studies 1-6 and 10):

Study 1. As with our pre-registered accuracy measure (average absolute error), there was a significant interaction between the right-skewed distribution and the Sliders, b = -0.06, SE = 0.02, p < .001, such that the Sliders generated significantly smaller error when the true distribution was right-skewed (b = -0.03, SE = 0.01, p = .005), but generated significantly greater error when the

true distribution was left-skewed or symmetric (b = 0.03, SE = 0.01, p = .003). With all distributions combined, there was not a significant main effect of method, b = 0.01, SE = 0.01, p = .402. With both methods combined, there was a significant main effect of the left-skewed distribution (b = 0.05, SE = 0.01, p < .001) and a significant main effect of the right-skewed distribution (b = 0.04, SE = 0.01, p < .001), suggesting that the KS distance was the smallest when the true distribution was symmetric.

Study 2. As with our pre-registered primary measure, we found a significant interaction between distribution shape and method, b = -0.03, SE = 0.01, p < .001, such that the Sliders generated significantly greater error when the true distribution was left-skewed, b = 0.02, SE = 0.01, p < .001, but significantly lower error when the true distribution was right-skewed, b = -0.01, SE = 0.01, p = .029. We also found the same order effect as with our primary measure. An interaction between the method and the order (b = -0.03, SE = 0.01, p < .001) reveals that when participants completed the Slider task before the Distribution Builder task, Slider elicited greater error than Distribution Builder, b = 0.02, SE = 0.01, p = .001, but this difference was attenuated when participants completed the Slider task after the Distribution Builder task, b = -0.01, SE = 0.01, p = .160.

Study 3. As with our primary measure, Sliders elicited responses with greater error with both left-skewed and right-skewed distributions combined, b = 0.01, SE = 0.003, p = .040. The interaction between method and order was not significant, albeit in the same direction as with the primary measure and with Study 2 results, b = -0.01, SE = 0.01, p = .088. That is, when participants first completed the Slider task, Sliders led to significantly greater error than Distribution Builder, b = 0.01, SE = 0.01, p = .003. This difference was nevertheless attenuated when participants first completed a Distribution Builder task and then a Slider task, b = 0.001, SE = 0.004, p = .823.

Study 4. As with our primary accuracy measure, we replicated the interaction in Study 1 with even partition conditions (b = -0.03, SE = 0.01, p = .020), such that Distribution Builder was more accurate when the shape was symmetric, but Sliders were more accurate when the shape was right-skewed. With uneven partition conditions, we directionally replicated the interaction effect with average absolute error (b = 0.02, SE = 0.01, p = .059) such that the interaction moves with how the shape *appeared* in the interface rather than the true shape of the underlying distribution. Overall, the three-way interaction was significant, b = 0.06, SE = 0.02, p = .004. Consistent with what we report in the paper, there was a main effect of right-skewed distribution (b = -0.01, SE = 0.004, p = .005), a main effect of uneven partition (b = 0.02, SE = 0.01, p = .001), but not a significant main effect of method (b = 0.002, SE = 0.01, p = .789).

Study 5. As with our primary measure, Sliders led to significantly greater error than Distribution Builder regardless of the distribution shape, b = 0.02, SE = 0.01, p = .001. The main effect of the distribution shape (b = -0.01, SE = 0.01, p = .250) and the interaction effect (b = 0.001, SE = 0.01, p = .915) were not significant.

Study 6. Consistent with our primary measure, Sliders led to significantly greater error than Distribution Builder overall, b = 0.01, SE = 0.004, p < .001. In addition, right-skewed distributions produced significantly smaller errors than left-skewed distributions, b = -0.01, SE = 0.003, p < .001. Whether the numbers were random or meaningful did not have a significant influence on the result, b = -0.001, SE = 0.004, p = .772. None of the interactions was significant (ps ≥ .179).

Study 10. Again, the results are consistent with our pre-registered measure. First, when we compare Distribution Builder and Regular Sliders, the interaction effect between method and distribution shape (b = -0.03, SE = 0.01, p = .045) suggests that when the normative distribution

was right-skewed, the two methods were not significantly different from each other (p = .910), but Regular Sliders generated larger KS-distance when the normative shape was left-skewed (p = .004). This pattern reversed when we compare Distribution Builder with Reverse Sliders. A significant interaction (b = 0.04, SE = 0.01, p = .017) reveals that when the normative distribution was right-skewed (and hence appeared left-skewed in the Reverse Sliders interface), Reverse Sliders led to significantly greater KS-distance (p = .018) but the condition difference was attenuated when the normative shape was left-skewed (p = .372).

Aggregate deviation in the full distribution (Study 7):

In Study 7, we asked participants to provide subjective probability distributions for four prediction questions. Because the correct answers for these questions were single numbers rather than full distributions, we cannot easily derive the KS distance. Instead, we derived the mass allocated each bin, multiplied that by the absolute distance between each bin to the correct bin, and then summed all these products. The resulting measure captures how much the full distribution participants produced deviated from the correct answer.

We replicated the results of our primary accuracy measures. Overall, elicited distributions in the Sliders condition had larger aggregate deviation than the Distribution Builder condition, b = 4.92, SE = 1.63, p = .003. This was qualified by a significant interaction (b = -6.98, SE = 3.11, p = .025) such that Sliders led to more deviation than Distribution Builder when the normative shape was symmetric (b = 8.58, SE = 2.11, p < .001) but the difference was attenuated when the normative shape was right-skewed (b = 1.60, SE = 2.38, p = .503). There was also a significant main effect of shape such that right-skewed distributions lead to less deviation than symmetric distributions, b = -13.81, SE = 1.55, p < .001.

Section 2.2: Comparisons to Intuitive Benchmarks

To facilitate an intuitive understanding of the means presented in Table 2 in the paper, we compare them to two benchmarks below: (1) in Table S2.2.1, the average absolute deviation generated by a completely uniform distribution (that is, by providing the same response in every bin) and (2) in Table S2.2.2., the average absolute deviation generated by a completely peaked distribution (that is, by allocating all the mass to modal bin in the normative distribution).

For ease of interpretation, we converted the raw means to percentages of error generated by the uniform distribution and the completely peaked distribution. For example, in the second row of Table S2.2.1, when the correct answer was left-skewed in Study 1, the Distribution Builder condition generated an average absolute error that was 82.5% of the error that we would observe if participants had instead just provided a uniform distribution (by providing the same response in every bin). That relative error increased to 88.0% in the Sliders condition. In the last column, this relative-to-uniform-benchmark error increased by 5.5 percentage points in the Sliders condition than in the Distribution Builder condition.

Table S2.2.1. Comparing Studies 1-6 Results Against a Completely Uniform Distribution

	Distribution Builder	Slider	Benchmark: Uniform Distribution			
	M (SD)	*M (SD)*	Benchmark AAE	DB %	Slider %	Slider%-DB%
Study 1						
Right-skewed	3.43 (1.63)	3.16 (1.50)	4.16	82.5%	76.0%	-6.5%
Left-skewed	3.43 (1.75)	3.66 (2.24)	4.16	82.5%	88.0%	5.5%
Symmetric	2.85 (1.50)	3.16 (1.93)	4.73	60.3%	66.8%	6.6%
Study 2						
Right-skewed	2.51 (1.18)	2.45 (1.22)	2.78	90.3%	88.1%	-2.2%
Left-skewed	**2.68 (1.40)**	2.97 (1.59)	2.78	96.4%	106.8%	10.4%
Study 3						
Right-skewed	**2.42 (1.15)**	2.59 (1.21)	2.78	87.1%	93.2%	6.1%
Left-skewed	**2.60 (1.37)**	2.75 (1.46)	2.78	93.5%	98.9%	5.4%
Study 4						
Actually Right-skewed / Looks Right-skewed	3.51 (2.07)	3.37 (1.81)	4.16	84.4%	81.0%	-3.4%
Actually Symmetric / Looks Symmetric	3.44 (1.93)	3.65 (2.08)	4.73	72.7%	77.2%	4.4%
Actually Right-skewed / Looks Symmetric	3.40 (1.94)	3.51 (1.70)	4.16	81.7%	84.4%	2.6%
Actually Symmetric / Looks Right-skewed	3.59 (1.93)	3.53 (1.94)	4.41	81.4%	80.0%	-1.4%
Study 5						
Right-skewed	**1.59 (0.78)**	1.73 (0.95)	1.60	99.4%	108.1%	8.7%
Symmetric	1.68 (0.85)	1.80 (1.04)	0.80	210.0%	225.0%	15.0%
Study 6						
Right-skewed / Meaningful Numbers	**2.42 (1.16)**	2.61 (1.27)	2.78	87.1%	93.9%	6.8%
Left-skewed / Meaningful Numbers	**2.55 (1.24)**	2.86 (1.45)	2.78	91.7%	102.9%	11.2%
Right-skewed / Random Numbers	**2.47 (1.28)**	2.73 (1.36)	2.78	88.8%	98.2%	9.4%
Left-skewed / Random Numbers	**2.57 (1.33)**	2.85 (1.57)	2.78	92.4%	102.5%	10.1%

Note.

1. Within each row, **boldface** indicates that this method generated significantly smaller AAE and was significantly more accurate.
2. The right hand side of the table presents the AAE of a completely uniform distribution as a benchmark. The two columns labeled "DB%" and "Slider%" convert the AAE generated by Distribution Builder and Sliders to a % of the error of uniform distribution. The last column labeled "Slider%-DB%" presents the difference between the two methods as a % of the uniform distribution error.

The same interpretation applies to comparisons with the peaked distribution. For instance, in the second row of Table S2.2.2, when the correct answer was left-skewed in Study 1, the Distribution Builder condition generated an average absolute error that was 37.5% of the error that we would observe if participants had put all of the mass into a single bin that matches the correct peak of the normative distribution. That relative error increased to 40.0% in the Sliders condition. So, in the last column, this relative-to-peaked-benchmark error increased by 2.5 percentage points in the Slider condition than in the Distribution Builder condition.

Table S2.2.2. Comparing Studies 1-6 Results Against a Completely Peaked Distribution

	Distribution Builder	Slider	Benchmark: Peaked Distribution			
	M (SD)	*M (SD)*	Benchmark AAE	DB %	Slider %	Slider%-DB%
Study 1						
Right-skewed	3.43 (1.63)	3.16 (1.50)	9.71	35.3%	32.5%	-2.8%
Left-skewed	3.43 (1.75)	3.66 (2.24)	9.14	37.5%	40.0%	2.5%
Symmetric	2.85 (1.50)	3.16 (1.93)	9.14	31.2%	34.6%	3.4%
Study 2						
Right-skewed	2.51 (1.18)	2.45 (1.22)	9.75	25.7%	25.1%	-0.6%
Left-skewed	**2.68 (1.40)**	2.97 (1.59)	9.75	27.5%	30.5%	3.0%
Study 3						
Right-skewed	**2.42 (1.15)**	2.59 (1.21)	9.75	24.8%	26.6%	1.7%
Left-skewed	**2.60 (1.37)**	2.75 (1.46)	9.75	26.7%	28.2%	1.5%
Study 4						
Actually Right-skewed / Looks Right-skewed	3.51 (2.07)	3.37 (1.81)	9.71	36.1%	34.7%	-1.4%
Actually Symmetric / Looks Symmetric	3.44 (1.93)	3.65 (2.08)	9.14	37.6%	39.9%	2.3%
Actually Right-skewed / Looks Symmetric	3.40 (1.94)	3.51 (1.70)	9.14	37.2%	38.4%	1.2%
Actually Symmetric / Looks Right-skewed	3.59 (1.93)	3.53 (1.94)	9.71	37.0%	36.4%	-0.6%
Study 5						
Right-skewed	**1.59 (0.78)**	1.73 (0.95)	5.20	30.6%	33.3%	2.7%
Symmetric	1.68 (0.85)	1.80 (1.04)	5.60	30.0%	32.1%	2.1%
Study 6 (All Tasks Combined)						
Right-skewed / Meaningful Numbers	**2.42 (1.16)**	2.61 (1.27)	9.75	24.8%	26.8%	1.9%
Left-skewed / Meaningful Numbers	**2.55 (1.24)**	2.86 (1.45)	9.75	26.2%	29.3%	3.2%
Right-skewed / Random Numbers	**2.47 (1.28)**	2.73 (1.36)	9.75	25.3%	28.0%	2.7%
Left-skewed / Random Numbers	**2.57 (1.33)**	2.85 (1.57)	9.75	26.4%	29.2%	2.9%

Note.

1. Within each row, **boldface** indicates that this method generated significantly smaller AAE and was significantly more accurate.
2. The right hand side of the table presents as a benchmark the AAE of a completely peaked distribution, where participants put all the mass into the correct modal bin. The two columns labeled "DB%" and "Slider%" convert the AAE generated by Distribution Builder and Sliders to a % of the error of a peaked distribution. The last column labeled "Slider%-DB%" presents the difference between the two methods as a % of the peaked distribution error.

Section 2.3: Results of Additional Order and Allocation Measures

In Table 3 in the paper, we presented the results of several exploratory measures (percentage of participants who started from the first bin; percentage of mass allocated to the first bin). As discussed in the paper, the results of these measures suggest that participants using Sliders were more likely to start from the first bin and, as a result, allocate more mass to the first few bins in the interface, which led them to produce overly right-skewed responses. The measures reported in Table 3 were among a series of other exploratory measures, which we report in Table S2.3.1 and Table S2.3.2.

Table S2.3.1 presents the results of all mechanism measures mentioned as “exploratory” in our pre-registrations. In addition to the finding that Sliders tend to prompt participants to start from the first few bins, the other measures did not reveal a systematic pattern.

Table S2.3.2 presents two measures regarding the modal bin in the distribution participants constructed. We derived for each elicited distribution the bin in which participants allocated the most mass and looked at the percentage of participants who started from that bin (the first measure) and the percentage of mass allocated to that bin (the second measure). There was no consistent pattern across studies, but in the cases where the two methods differed significantly from each other, Distribution Builder users were more likely to start from where they ended up allocating the most mass to, while Slider users tend to allocate more mass to their modal bin.

Table S2.3.1. Additional Exploratory Measures for Studies 4-7 and 10

Study, Measure, and Distribution	Distribution Builder	Sliders	Effect of the Sliders Condition (vs. the Distribution Builder Condition)
Study 4			
% of Participants Who Started From the First Two Buckets			
Actually Right-skewed / Looks Right-skewed	66.92%	**80.80%**	$b = 0.14$, $SE = 0.03$, $p < .001$
Actually Symmetric / Looks Symmetric	62.81%	**73.19%**	$b = 0.10$, $SE = 0.03$, $p = .002$
Actually Right-skewed / Looks Symmetric	64.00%	**76.13%**	$b = 0.12$, $SE = 0.03$, $p < .001$
Actually Symmetric / Looks Right-skewed	62.63%	**77.35%**	$b = 0.15$, $SE = 0.03$, $p < .001$
% of Participants Who Started From the Middle Three Buckets			
Actually Right-skewed / Looks Right-skewed	**17.42%**	5.60%	$b = -0.12$, $SE = 0.02$, $p < .001$
Actually Symmetric / Looks Symmetric	**20.10%**	11.53%	$b = -0.09$, $SE = 0.03$, $p = .001$
Actually Right-skewed / Looks Symmetric	**17.87%**	7.40%	$b = -0.10$, $SE = 0.02$, $p < .001$
Actually Symmetric / Looks Right-skewed	**18.82%**	12.05%	$b = -0.07$, $SE = 0.03$, $p = .008$
% of Participants Who Started From the Last Two Buckets			
Actually Right-skewed / Looks Right-skewed	15.66%	10.93%	$b = -0.05$, $SE = 0.02$, $p = .054$
Actually Symmetric / Looks Symmetric	**17.09%**	12.06%	$b = -0.05$, $SE = 0.03$, $p = .049$
Actually Right-skewed / Looks Symmetric	**18.13%**	12.89%	$b = -0.05$, $SE = 0.03$, $p = .041$
Actually Symmetric / Looks Right-skewed	**18.55%**	8.67%	$b = -0.10$, $SE = 0.02$, $p < .001$
Study 5			
% of Participants Who Started From the Middle Bucket			
Right-skewed	**5.24%**	2.17%	$b = -0.03$, $SE = 0.01$, $p = .027$
Symmetric	**7.14%**	2.29%	$b = -0.05$, $SE = 0.02$, $p = .001$
% of Participants Who Started From the Last Bucket			
Right-skewed	23.04%	17.89%	$b = -0.05$, $SE = 0.03$, $p = .081$
Symmetric	27.04%	21.37%	$b = -0.06$, $SE = 0.03$, $p = .064$
Study 6			
% of Participants Who Started From the Middle Two Buckets			
Right-skewed / Meaningful Numbers	**10.82%**	5.38%	$b = -0.05$, $SE = 0.02$, $p = .003$
Left-skewed / Meaningful Numbers	11.56%	8.21%	$b = -0.04$, $SE = 0.02$, $p = .056$
Right-skewed / Random Numbers	**12.63%**	5.66%	$b = -0.07$, $SE = 0.02$, $p < .001$
Left-skewed / Random Numbers	**15.53%**	7.56%	$b = -0.08$, $SE = 0.02$, $p < .001$
% of Participants Who Started From the Last Bucket			
Right-skewed / Meaningful Numbers	8.98%	7.96%	$b = -0.01$, $SE = 0.02$, $p = .530$
Left-skewed / Meaningful Numbers	**13.79%**	8.21%	$b = -0.05$, $SE = 0.02$, $p = .007$
Right-skewed / Random Numbers	6.85%	6.50%	$b = -0.002$, $SE = 0.02$, $p = .905$
Left-skewed / Random Numbers	12.55%	11.97%	$b = -0.01$, $SE = 0.02$, $p = .670$
Absolute Difference Between the Mass Allocated to the First Half vs. the Second Half			
Right-skewed / Meaningful Numbers	16.37	16.61	$b = 0.21$, $SE = 0.68$, $p = .752$
Left-skewed / Meaningful Numbers	14.75	14.82	$b = 0.08$, $SE = 0.66$, $p = .900$
Right-skewed / Random Numbers	15.46	15.70	$b = 0.23$, $SE = 0.70$, $p = .747$
Left-skewed / Random Numbers	14.46	14.80	$b = 0.39$, $SE = 0.69$, $p = .574$
Study 7			
% of Participants Who Started From the Middle Bucket			
Right-skewed	**6.53%**	4.45%	$b = -0.02$, $SE = 0.01$, $p = .013$
Symmetric	**25.44%**	15.20%	$b = -0.10$, $SE = 0.01$, $p < .001$
% of Participants Who Started From the Last Bucket			
Right-skewed	0.90%	1.27%	$b = 0.004$, $SE = 0.004$, $p = .321$
Symmetric	3.44%	2.87%	$b = -0.01$, $SE = 0.01$, $p = .407$
Absolute Difference Between the Mass Allocated to the First Half vs. the Second Half			
Right-skewed	80.89	81.03	$b = 0.07$, $SE = 1.00$, $p = .944$
Symmetric	56.33	56.99	$b = 0.66$, $SE = 1.23$, $p = .594$
Study 10			
% of Participants Who Started From the Middle Bucket			
Right-skewed / Regular Sliders	5.23%	2.71%	$b = -0.03$, $SE = 0.02$, $p = .120$
Right-skewed / Reverse Sliders	5.23%	2.30%	$b = -0.03$, $SE = 0.02$, $p = .061$
Left-skewed / Regular Sliders	3.93%	2.93%	$b = -0.01$, $SE = 0.01$, $p = .496$
Left-skewed / Reverse Sliders	3.93%	1.38%	$b = -0.03$, $SE = 0.01$, $p = .055$
% of Participants Who Started From the Last Bucket			
Right-skewed / Regular Sliders	6.62%	9.15%	$b = 0.03$, $SE = 0.02$, $p = .258$
Right-skewed / Reverse Sliders	6.62%	**15.13%**	$b = 0.09$, $SE = 0.03$, $p = .001$
Left-skewed / Regular Sliders	13.77%	13.03%	$b = -0.01$, $SE = 0.03$, $p = .788$
Left-skewed / Reverse Sliders	13.77%	17.65%	$b = 0.04$, $SE = 0.03$, $p = .194$
Absolute Difference Between the Mass Allocated to the First Half vs. the Second Half			
Right-skewed / Regular Sliders	21.56	22.00	$b = 0.45$, $SE = 1.01$, $p = .659$
Right-skewed / Reverse Sliders	21.56	20.50	$b = -1.06$, $SE = 1.01$, $p = .294$
Left-skewed / Regular Sliders	22.68	21.13	$b = -1.55$, $SE = 1.03$, $p = .132$
Left-skewed / Reverse Sliders	22.68	23.90	$b = 1.22$, $SE = 1.02$, $p = .232$

Note. Within each row, **boldface** indicates that this percentage is significantly higher than the adjacent percentage.

Table S2.3.2. Additional Measures Regarding Modal Bucket for Studies 1-10

		Distribution Builder	Sliders	Effect of the Sliders Condition (vs. the Distribution Builder Condition)
% of Participants Who Started from the Modal Bucket in their own Distribution				
Study 2	Right-skewed	**13.69%**	9.47%	b = -0.04, SE = 0.02, p = .040
	Left-skewed	**16.43%**	11.50%	b = -0.05, SE = 0.02, p = .027
Study 3	Right-skewed	**13.75%**	9.52%	b = -0.04, SE = 0.02, p = .009
	Left-skewed	**14.43%**	10.95%	b = -0.03, SE = 0.02, p = .037
Study 4	Actually Right-skewed / Looks Right-skewed	17.42%	18.67%	b = 0.01, SE = 0.03, p = .654
	Actually Symmetric / Looks Symmetric	16.83%	13.94%	b = -0.03, SE = 0.03, p = .267
	Actually Right-skewed / Looks Symmetric	**13.60%**	8.35%	b = -0.05, SE = 0.02, p = .018
	Actually Symmetric / Looks Right-skewed	14.78%	14.94%	b = 0.002, SE = 0.03, p = .951
Study 5	Right-skewed	22.51%	23.04%	b = 0.01, SE = 0.03, p = .865
	Symmetric	15.05%	15.52%	b = 0.01, SE = 0.03, p = .855
Study 6	Right-skewed / Meaningful Numbers	**17.14%**	12.26%	b = -0.05, SE = 0.02, p = .036
	Left-skewed / Meaningful Numbers	14.20%	11.02%	b = -0.03, SE = 0.02, p = .130
	Right-skewed / Random Numbers	**26.55%**	16.14%	b = -0.10, SE = 0.03, p < .001
	Left-skewed / Random Numbers	**22.98%**	13.03%	b = -0.10, SE = 0.02, p < .001
Study 7	Right-skewed	**54.29%**	43.33%	b = -0.11, SE = 0.02, p < .001
	Symmetric	**51.21%**	31.93%	b = -0.19, SE = 0.02, p < .001
Study 8	-	**46.39%**	32.76%	b = -0.14, SE = 0.03, p < .001
Study 9	-	**55.56%**	46.42%	b = -0.09, SE = 0.02, p < .001
Study 10	Right-skewed / Regular Sliders	23.34%	21.69%	b = -0.02, SE = 0.03, p = .634
	Right-skewed / Reverse Sliders	**23.34%**	14.80%	b = -0.09, SE = 0.03, p = .008
	Left-skewed / Regular Sliders	16.39%	17.92%	b = 0.02, SE = 0.03, p = .618
	Left-skewed / Reverse Sliders	16.39%	21.11%	b = 0.05, SE = 0.03, p = .141
% of Mass Allocated to the Modal Bucket in their own Distribution				
Study 1	Right-skewed	32.34%	33.03%	b = 0.01, SE = 0.01, p = .461
	Left-skewed	34.57%	35.40%	b = 0.01, SE = 0.01, p = .418
	Symmetric	33.11%	34.00%	b = 0.01, SE = 0.01, p = .322
Study 2	Right-skewed	25.92%	**26.94%**	b = 0.01, SE = 0.01, p = .039
	Left-skewed	25.34%	**26.40%**	b = 0.01, SE = 0.01, p = .042
Study 3	Right-skewed	25.18%	**27.11%**	b = 0.02, SE = 0.004, p < .001
	Left-skewed	25.38%	26.03%	b = 0.01, SE = 0.004, p = .085
Study 4	Actually Right-skewed / Looks Right-skewed	33.90%	32.74%	b = -0.01, SE = 0.01, p = .181
	Actually Symmetric / Looks Symmetric	33.48%	33.32%	b = -0.002, SE = 0.01, p = .852
	Actually Right-skewed / Looks Symmetric	31.56%	30.73%	b = -0.01, SE = 0.01, p = .250
	Actually Symmetric / Looks Right-skewed	32.64%	31.91%	b = -0.01, SE = 0.01, p = .342
Study 5	Right-skewed	42.59%	**44.59%**	b = 0.02, SE = 0.01, p = .020
	Symmetric	38.01%	38.73%	b = 0.01, SE = 0.01, p = .351
Study 6	Right-skewed / Meaningful Numbers	25.21%	**26.16%**	b = 0.01, SE = 0.004, p = .023
	Left-skewed / Meaningful Numbers	25.27%	25.89%	b = 0.01, SE = 0.01, p = .219
	Right-skewed / Random Numbers	24.42%	**25.55%**	b = 0.01, SE = 0.01, p = .041
	Left-skewed / Random Numbers	24.32%	25.32%	b = 0.01, SE = 0.01, p = .068
Study 7	Right-skewed	54.56%	55.24%	b = -0.01, SE = 0.01, p = .391
	Symmetric	51.21%	50.69%	b = 0.01, SE = 0.01, p = .337
Study 8	-	41.60%	43.35%	b = 0.02, SE = 0.01, p = .116
Study 9	-	44.22%	**46.01%**	b = 0.02, SE = 0.01, p = .010
Study 10	Right-skewed / Regular Sliders	31.05%	32.12%	b = 0.01, SE = 0.01, p = .183
	Right-skewed / Reverse Sliders	31.05%	30.99%	b < -0.001, SE = 0.01, p = .927
	Left-skewed / Regular Sliders	31.92%	32.23%	b = 0.003, SE = 0.01, p = .696
	Left-skewed / Reverse Sliders	31.92%	31.18%	b = -0.01, SE = 0.01, p = .315

Note. Within each row, **boldface** indicates that this percentage is significantly higher than the adjacent percentage.

Section 2.4: Moments of Elicited Distributions

Table S2.4 presents the results of all moment measures mentioned as "exploratory" in our pre-registrations, by study.

Table S2.4. Exploratory Moments Measures for Studies 1, 4-7, and 10

		Distribution Builder		Sliders		Effect of the Sliders condition
Study, Measure, and Distribution	True Distribution	*Mean*	*Median*	*Mean*	*Median*	(vs. Distribution Builder condition)
Study 1						
Distribution Mean						
Right-skewed	21.70	24.09	23.80	**23.10**	23.20	b = -0.99, SE = 0.24, p < .001
Left-skewed	34.10	31.83	32.40	31.94	32.50	b = 0.11, SE = 0.31, p = .722
Symmetric	28.00	27.76	28.00	27.37	27.50	b = -0.39, SE = 0.24, p = .103
Distribution Standard Deviation						
Right-skewed	7.06	6.72	6.70	6.58	6.58	b = -0.14, SE = 0.15, p = .350
Left-skewed	6.87	6.67	6.63	6.52	6.60	b = -0.15, SE = 0.15, p = .318
Symmetric	6.31	6.41	6.29	6.35	6.33	b = -0.07, SE = 0.15, p = .664
Distribution Skewness						
Right-skewed	0.62	0.25	0.22	0.34	0.33	b = 0.08, SE = 0.05, p = .065
Left-skewed	-0.66	-0.31	-0.34	-0.36	-0.38	b = -0.05, SE = 0.04, p = .177
Symmetric	0.00	0.02	0.00	0.05	0.01	b = 0.03, SE = 0.04, p = .417
Distribution Kurtosis						
Right-skewed	2.42	2.46	2.36	2.64	2.35	b = 0.18, SE = 0.13, p = .188
Left-skewed	2.51	2.57	2.41	2.62	2.48	b = 0.05, SE = 0.07, p = .442
Symmetric	2.75	2.46	2.41	2.50	2.42	b = 0.04, SE = 0.06, p = .521

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

		Distribution Builder		Sliders		Effect of the Sliders condition
Study, Measure, and Distribution	True Distribution	*Mean*	*Median*	*Mean*	*Median*	(vs. Distribution Builder condition)
Study 4						
Distribution Mean						
Actually Right-skewed / Looks Right-skewed	21.70	23.83	23.10	23.44	23.00	b = -0.39, SE = 0.24, p = .107
Actually Symmetric / Looks Symmetric	28.00	**26.07**	26.20	25.46	25.50	b = -0.61, SE = 0.26, p = .019
Actually Right-skewed / Looks Symmetric	22.17	22.14	21.94	22.05	22.12	b = -0.09, SE = 0.20, p = .664
Actually Symmetric / Looks Right-skewed	28.00	**32.13**	32.37	31.58	32.01	b = -0.56, SE = 0.24, p = .022
Distribution Standard Deviation						
Actually Right-skewed / Looks Right-skewed	7.06	6.76	6.65	6.93	6.76	b = 0.17, SE = 0.14, p = .205
Actually Symmetric / Looks Symmetric	6.31	6.59	6.41	6.73	6.58	b = 0.14, SE = 0.14, p = .310
Actually Right-skewed / Looks Symmetric	7.20	7.41	7.23	7.42	7.31	b = 0.01, SE = 0.13, p = .950
Actually Symmetric / Looks Right-skewed	7.24	6.81	6.46	6.84	6.61	b = 0.03, SE = 0.14, p = .810
Distribution Skewness						
Actually Right-skewed / Looks Right-skewed	0.62	0.28	0.29	0.28	0.29	b = -0.001, SE = 0.03, p = .970
Actually Symmetric / Looks Symmetric	0.00	0.11	0.06	0.11	0.11	b = 0.01, SE = 0.03, p = .845
Actually Right-skewed / Looks Symmetric	0.75	0.60	0.59	0.56	0.57	b = -0.04, SE = 0.03, p = .186
Actually Symmetric / Looks Right-skewed	0.00	-0.18	-0.22	-0.15	-0.18	b = 0.02, SE = 0.03, p = .439
Distribution Kurtosis						
Actually Right-skewed / Looks Right-skewed	2.42	2.46	2.32	2.36	2.21	b = -0.10, SE = 0.06, p = .075
Actually Symmetric / Looks Symmetric	2.75	2.48	2.36	2.36	2.31	b = -0.11, SE = 0.06, p = .056
Actually Right-skewed / Looks Symmetric	2.91	2.66	2.48	2.60	2.41	b = -0.06, SE = 0.07, p = .378
Actually Symmetric / Looks Right-skewed	2.23	2.81	2.74	2.79	2.59	b = -0.02, SE = 0.06, p = .683

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

Study, Measure, and Distribution	True Distribution	Distribution Builder		Sliders		Effect of the Sliders condition (vs. Distribution Builder condition)
		Mean	*Median*	*Mean*	*Median*	
Study 5						
Distribution Mean						
Right-skewed	4.00	3.95	3.93	3.91	3.90	b = -0.04, SE = 0.02, p = .145
Symmetric	4.50	4.31	4.33	4.26	4.28	b = -0.05, SE = 0.03, p = .066
Distribution Standard Deviation						
Right-skewed	0.99	0.87	0.89	0.87	0.89	b = -0.004, SE = 0.01, p = .782
Symmetric	1.06	0.95	0.97	0.95	0.98	b = -0.004, SE = 0.01, p = .696
Distribution Skewness						
Right-skewed	0.78	0.78	0.75	0.81	0.76	b = 0.03, SE = 0.04, p = .466
Symmetric	0.00	0.28	0.24	0.31	0.26	b = 0.03, SE = 0.04, p = .414
Distribution Kurtosis						
Right-skewed	2.05	2.77	2.35	2.95	2.35	b = 0.18, SE = 0.12, p = .140
Symmetric	1.58	2.01	1.87	2.06	1.80	b = 0.05, SE = 0.06, p = .382

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

Study, Measure, and Distribution	True Distribution	Distribution Builder		Sliders		Effect of the Sliders condition (vs. Distribution Builder condition)
		Mean	*Median*	*Mean*	*Median*	
Study 6						
Distribution Mean						
Right-skewed (Heart rate)	64.92	66.93	66.88	66.78	66.80	b = -0.15, SE = 0.16, p = .354
Right-skewed (Mile time)	7.98	8.38	8.36	8.42	8.40	b = 0.04, SE = 0.04, p = .298
Left-skewed (Heart rate)	74.22	**71.83**	71.84	71.43	71.84	b = -0.40, SE = 0.20, p = .042
Left-skewed (Mile time)	10.02	9.53	9.56	9.50	9.52	b = -0.03, SE = 0.04, p = .471
Distribution Standard Deviation						
Right-skewed (Heart rate)	8.88	**8.47**	8.45	8.24	8.30	b = -0.24, SE = 0.10, p = .017
Right-skewed (Mile time)	1.85	1.79	1.81	1.78	1.81	b = -0.01, SE = 0.02, p = .500
Left-skewed (Heart rate)	8.99	8.68	8.77	8.58	8.71	b = -0.11, SE = 0.10, p = .299
Left-skewed (Mile time)	1.85	1.86	1.88	1.87	1.86	b = 0.01, SE = 0.02, p = .811

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

Study, Measure, and Distribution	True Distribution	Distribution Builder		Sliders		Effect of the Sliders condition (vs. Distribution Builder condition)
		Mean	*Median*	*Mean*	*Median*	
Study 7						
Distribution Mean						
Right-skewed (Weight)		177.10	176.60	176.84	176.60	b = -0.26, SE = 1.27, p = .839
Right-skewed (Age)		40.02	40.00	40.02	39.50	b < 0.001, SE = 0.27, p = 1.000
Right-skewed (Temperature)		88.66	88.03	88.27	87.75	b = -0.38, SE = 0.28, p = .170
Right-skewed (Walking Time)		29.35	27.50	28.86	27.18	b = -0.49, SE = 0.44, p = .265
Symmetric (Weight)		168.39	170.00	166.80	169.00	b = -1.60, SE = 1.31, p = .224
Symmetric (Age)		**37.37**	37.33	36.13	36.20	b = -1.24, SE = 0.32, p < .001
Symmetric (Temperature)		84.99	85.00	84.66	84.50	b = -0.33, SE = 0.43, p = .446
Symmetric (Walking Time)		23.73	22.93	23.97	23.05	b = 0.24, SE = 0.45, p = .589
Distribution Standard Deviation						
Right-skewed (Weight)		13.03	12.63	**14.01**	13.26	b = 0.97, SE = 0.38, p = .010
Right-skewed (Age)		3.57	3.44	3.59	3.52	b = 0.02, SE = 0.09, p = .798
Right-skewed (Temperature)		3.81	3.55	**4.03**	3.86	b = 0.22, SE = 0.11, p = .048
Right-skewed (Walking Time)		3.68	3.51	3.73	3.52	b = 0.04, SE = 0.13, p = .731
Symmetric (Weight)		13.54	13.33	13.70	13.63	b = 0.16, SE = 0.38, p = .669
Symmetric (Age)		3.75	3.61	**4.07**	3.92	b = 0.32, SE = 0.10, p = .001
Symmetric (Temperature)		4.35	4.14	**4.68**	4.52	b = 0.34, SE = 0.12, p = .004
Symmetric (Walking Time)		3.69	3.61	**4.05**	3.86	b = 0.36, SE = 0.11, p = .001

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

Study, Measure, and Distribution	True Distribution	Distribution Builder		Sliders		Effect of the Sliders condition (vs. Distribution Builder condition)
		Mean	*Median*	*Mean*	*Median*	
Study 10						
Distribution Mean						
Right-skewed / Regular Sliders	67.00	69.12	68.90	68.93	68.80	b = -0.19, SE = 0.25, p = .449
Right-skewed / Reverse Sliders	67.00	69.12	68.90	**69.90**	69.30	b = 0.78, SE = 0.28, p = .005
Left-skewed / Regular Sliders	79.00	**77.20**	77.50	76.41	76.80	b = -0.80, SE = 0.28, p = .005
Left-skewed / Reverse Sliders	79.00	77.20	77.50	77.68	77.80	b = 0.47, SE = 0.26, p = .068
Distribution Standard Deviation						
Right-skewed / Regular Sliders	7.49	7.27	7.14	7.31	7.34	b = 0.05, SE = 0.14, p = .726
Right-skewed / Reverse Sliders	7.49	7.27	7.14	7.39	7.52	b = 0.12, SE = 0.14, p = .371
Left-skewed / Regular Sliders	7.49	7.47	7.57	7.70	7.80	b = 0.23, SE = 0.14, p = .093
Left-skewed / Reverse Sliders	7.49	7.47	7.57	7.54	7.66	b = 0.07, SE = 0.13, p = .615

Note. Within each row, **boldface** indicates that this number is significantly higher than the adjacent number.

Supplement 3: Studies 1-7 Full Regression Results

In Studies 1-7, we pre-registered to conduct OLS regressions to determine whether Distribution Builder or Sliders elicited more accurate distributions and whether the effect depended on the shape of the distribution and any additional variable of interest (i.e., order in Studies 2 and 3, range partition in Study 4 and the meaningfulness of the numbers in Study 6). Specifically, in Studies 1, 5, and 7, we regressed the accuracy measure on (a) the method condition (contrast-coded), (b) the shape condition(s) (contrast-coded), and (c) the interaction(s) between the method condition and the shape condition(s). In Studies 2-4 and 6, we also included in the regression (d) the additional manipulation (contrast-coded; order in Studies 2 and 3, range partition in Study 4, and number perception in Study 6), (e,f) two-way interactions between the additional manipulation and the other two independent variables, and (g) the three-way interaction. In studies where participants provided more than one belief distribution (Studies 2-4, 6, and 7), the regressions clustered standard errors by participant. In studies with more than one item (Studies 3, 6, and 7), the regressions included fixed effects for item. In the paper, we report the means and pairwise comparisons (see Table 2 in the paper) and discuss any significant interactions. We report below the full regression results.

Study 1

To examine the effects of elicitation method and distribution shape on belief distribution accuracy, we regressed average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), (3) the left-skewed condition (contrast-coded), (4) the interaction between (1) and (2), and (5) the interaction between (1) and (3).

Table S3.1 presents the results. We found a significant interaction between the Sliders condition and the right-skewed condition, b = -0.58, SE = 0.25, p = .019. Specifically, when the true distribution was right-skewed, Sliders produced directionally smaller error (and thus higher accuracy) than Distribution Builder, b = -0.27, SE = 0.15, p = .079. But when the true distribution was left-skewed or symmetric, Sliders produced significantly greater error than Distribution Builder, b = 0.31, SE = 0.13, p = .017. With all distributions combined, the main effect of the elicitation method was not significant, b = 0.09, SE = 0.10, p = .360. With both methods combined, there was a significant main effect of the right-skewed distribution (b = 0.29, SE = 0.12, p = .019) and of the left-skewed distribution (b = 0.54, SE = 0.12, p < .001), such that the average absolute error was the lowest when the shape of the true distribution was symmetric (symmetric distribution: M = 3.00, SD = 1.73; right-skewed distribution: M = 3.30, SD = 1.58; left-skewed distribution: M = 3.56, SD = 2.04).

Table S3.1. Study 1 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Right-skewed	3.43	1.63	3.16	1.50	$t(418) = 1.76, p = .079$
Left-skewed	3.43	1.75	3.66	2.24	$t(428) = -1.19, p = .234$
Symmetric	2.85	1.50	3.16	1.93	$t(424) = -1.86, p = .064$

In sum, Study 1 suggests that Distribution Builder attained somewhat higher accuracy when the true distribution was left-skewed or symmetric, while Sliders attained somewhat higher accuracy for right-skewed distributions.

Study 2

In Study 2, all participants completed the memorize-recall task twice, once with a Distribution Builder interface, and once with a Sliders interface. We counterbalanced the order in which they used the two methods. As in Study 1, we calculated the average absolute error as the measure of individual-level accuracy. To test whether the accuracy depends on the elicitation method, the distribution shape, and the order, we regressed the average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), (3) the Distribution Builder first condition (contrast-coded), (4) the interaction between (1) and (2), (5) the interaction between (1) and (3), (6) the interaction between (2) and (3), and (7) the three-way interaction between (1), (2) and (3). We clustered standard errors by participant.

Table S3.2 presents the results. A main effect of method reveals that, overall, Sliders led to significantly greater error than Distribution Builder, $b = 0.12$, $SE = 0.05$, $p = .009$. We replicated the interaction effect with distribution shape in Study 1, $b = -0.38$, $SE = 0.09$, $p < .001$. Further, we found a significant interaction between method and order, $b = -0.55$, $SE = 0.09$, $p < .001$, suggesting that participants improved over time and generated distributions with less error in the second task. Specifically, when participants completed the first task using Sliders (and the second task using Distribution Builder), Sliders led to significantly greater error than Distribution Builder, $b = 0.39$, $SE = 0.07$, $p < .001$. However, when participants completed the first task using Distribution Builder (and the second task using Sliders), Sliders led to significantly smaller error than Distribution Builder, $b = -0.14$, $SE = 0.06$, $p = .020$. We replicated this interaction effect with order in Study 3.

Overall, a right-skewed distribution led to smaller error than a left-skewed distribution, $b = -0.34$, $SE = 0.07$, $p < .001$. The main effect of order was not significant ($b = 0.11$, $SE = 0.07$, p

= .143), nor was the interaction between distribution shape and order ($b = 0.20$, $SE = 0.15$, $p = .161$) or the three-way interaction ($b = 0.23$, $SE = 0.19$, $p = .216$).

Table S3.2. Study 2 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Right-skewed (All orders combined)	2.51	1.18	2.45	1.22	$t(966) = 0.77$, $p = .443$
Right-skewed (Distribution Builder first)	2.72	1.28	**2.44**	1.17	$t(469) = 2.47$, $p = .014$
Right-skewed (Sliders first)	2.30	1.04	2.45	1.26	$t(495) = -1.44$, $p = .149$
Left-skewed (All orders combined)	**2.68**	1.40	2.97	1.59	$t(972) = -3.03$, $p = .003$
Left-skewed (Distribution Builder first)	2.83	1.46	2.81	1.47	$t(515) = 0.16$, $p = .877$
Left-skewed (Sliders first)	**2.50**	1.31	3.14	1.71	$t(455) = -4.51$, $p < .001$

Note. Within each row, **boldface** indicates that this method generates significantly smaller AAE and is significantly more accurate.

Study 3

Study 3 had a similar design as Study 2 with the only difference being that we used different stimuli across the two trials and manipulated distribution shape (right-skewed or left-skewed) across the two trials orthogonally.

To test whether accuracy depends on the elicitation method, the distribution shape, and the order, we ran the same OLS regression as in Study 2. That is, we regressed the average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), (3) the Distribution Builder first condition (contrast-coded), (4) the interaction between (1) and (2), (5) the interaction between (1) and (3), (6) the interaction between (2) and (3), and (7) the three-way interaction between (1), (2) and (3). We clustered standard errors by participant and, because we used different stimuli across the two trials, included fixed effects for stimuli.

Table S3.3 presents the results. Replicating Study 2, we found a main effect of method such that Sliders generated greater error than Distribution Builder in general, $b = 0.15$, $SE = 0.04$, $p < .001$. Different from Study 2, this was not qualified by an interaction with the distribution shape,

b = 0.02, SE = 0.09, p = .803, such that Distribution Builder yielded more accurate results for both right-skewed and left-skewed distributions (see Table S3.3). We again found a significant interaction between method and order, b = -0.33, SE = 0.08, p < .001. That is, when participants first completed the Sliders interface (and then the Distribution Builder interface), Sliders led to significantly greater error than Distribution Builder, b = 0.34, SE = 0.05, p < .001. However, the difference between the two methods was attenuated when participants first completed the Distribution Builder and then the Sliders, b = -0.02, SE = 0.05, p = .704.

Table S3.3. Study 3 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Right-skewed (All orders combined)	**2.42**	1.15	2.59	1.21	$t(1,575) = -2.96, p = .003$
Right-skewed (Distribution Builder first)	2.56	1.19	2.51	1.16	$t(750) = 0.57, p = .566$
Right-skewed (Sliders first)	**2.29**	1.10	2.67	1.25	$t(823) = -4.61, p < .001$
Left-skewed (All orders combined)	**2.60**	1.37	2.75	1.46	$t(1,585) = -2.02, p = .043$
Left-skewed (Distribution Builder first)	2.59	1.29	2.61	1.51	$t(815) = -0.21, p = .832$
Left-skewed (Sliders first)	**2.61**	1.46	2.87	1.41	$t(768) = -2.53, p = .012$

Note. Within each row, **boldface** indicates that this method generates significantly smaller AAE and is significantly more accurate.

In secondary analysis, we examined whether the order effect – that the difference between Distribution Builder and Sliders was attenuated when participants first completed the Distribution Builder task and then the Sliders task – depends on the distribution shape for each method. Within each order condition, we regressed the average absolute error on the method condition (contrast-coded), the Distribution Builder shape condition (contrast-coded), the Sliders shape condition (contrast-coded), all the two-way interactions, and the three-way interaction. None of the interactions were significant. Nevertheless, as Table S3.3.2 shows, when participants completed the Sliders task after the Distribution Builder task, they only became directionally more accurate when the true distribution in the Slider task was right-skewed. This is again consistent with our

finding that the only case when Sliders elicited more accurate responses was when the true distribution had a right-skewed shape.

Table S3.3.2. Study 3 Results by Order and Distribution Shape

Order: Distribution Builder first, Sliders second				
Shape of the true distribution		Distribution Builder	Sliders	Which method is directionally
Distribution Builder	*Sliders*	*M (SD)*	*M (SD)*	more accurate?
Left-skewed	*Left-skewed*	2.50 (1.30)	2.63 (1.47)	Distribution Builder
Left-skewed	*Right-skewed*	2.68 (1.28)	2.55 (1.13)	Sliders
Right-skewed	*Left-skewed*	2.60 (1.24)	2.60 (1.54)	Distribution Builder
Right-skewed	*Right-skewed*	2.52 (1.14)	2.47 (1.21)	Sliders
Order: Sliders first, Distribution Builder second				
Shape of the true distribution		Distribution Builder	Sliders	Which method is directionally
Sliders	*Distribution Builder*	*M (SD)*	*M (SD)*	more accurate?
Left-skewed	*Left-skewed*	**2.61 (1.39)**	2.91 (1.43)	Distribution Builder
Right-skewed	*Left-skewed*	2.61 (1.54)	2.75 (1.38)	Distribution Builder
Left-skewed	*Right-skewed*	**2.38 (1.20)**	2.84 (1.40)	Distribution Builder
Right-skewed	*Right-skewed*	**2.21 (0.99)**	2.61 (1.12)	Distribution Builder

Note. Within each row, **boldface** indicates that this method generates significantly smaller AAE and is significantly more accurate.

Study 4

As in previous studies, we calculated the average absolute error as the measure of individual-level accuracy. To test whether the accuracy of the recalled distribution depends on the elicitation method, the distribution shape, and the range partition, we regressed the average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), (3) the uneven partition condition (contrast-coded), (4) the interaction between (1) and (2), (5) the interaction between (1) and (3), (6) the interaction between (2) and (3), and (7) the three-way interaction between (1), (2) and (3). We clustered standard errors by participant.

Table S3.4 summarizes the results. First, when the ranges were evenly partitioned such that how the distribution *appears* matched its true underlying shape (as in Study 1), we replicated the interaction effect in Study 1. Distribution Builder led to directionally smaller error in the

symmetric condition but Sliders led to directionally smaller error in the right-skewed condition, b = -0.45, SE = 0.20, p = .025 (top two rows in Table S3.4).

Was this interaction mostly driven by the true underlying shape or how the distribution shape appears? Results in the uneven partition conditions suggest it was the latter. As shown in the bottom two rows in Table S3.4, Distribution Builder was directionally more accurate when the distribution *appeared* symmetric, whereas the Sliders were directionally more accurate when the distribution *appeared* right-skewed. Thus, the Sliders × right-skewed interaction for uneven partition conditions followed the direction of apparent shape (b = 0.31, SE = 0.19, p = .106). The three-way interaction between elicitation method, distribution shape, and range partition was significant, b = 0.76, SE = 0.29, p = .009.

Overall, inherently right-skewed distributions led to significantly smaller errors than inherently symmetric distributions (b = -0.11, SE = 0.05, p = .032), and even partition led to significantly smaller errors than uneven partition (b = -0.19, SE = 0.07, p = .004). With all distribution conditions combined, the two elicitation methods did not differ in their accuracy, b = 0.02, SE = 0.08, p = .781.

Table S3.4. Study 4 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Actually Right-skewed / Looks Right-skewed	3.51	2.07	3.37	1.81	$t(769) = 0.98, p = .326$
Actually Symmetric / Looks Symmetric	3.44	1.93	3.65	2.08	$t(769) = -1.43, p = .154$
Actually Right-skewed / Looks Symmetric	3.40	1.94	3.51	1.70	$t(792) = -0.88, p = .378$
Actually Symmetric / Looks Right-skewed	3.59	1.93	3.53	1.94	$t(785) = 0.50, p = .621$

Taken together, we directionally replicated the results in Studies 1-3 that Distribution Builder attained higher accuracy for symmetric distribution and Sliders attained higher accuracy for right-skewed distribution. Furthermore, Study 4 suggests that this interaction is primarily driven by what the response distribution looks like rather than the true shape of the underlying distribution. When the distribution appears to be a different shape from its true underlying shape in the uneven

partition conditions, the direction of the interaction followed the apparent shapes in that Sliders were slightly more accurate for distribution that appeared right-skewed and Distribution Builder was slightly more accurate for distribution that appeared symmetric.

Study 5

Our dependent measure was average absolute error, calculated the same way as in previous studies. To examine whether the accuracy of reproduced distribution differs by the elicitation method and the distribution shape, we regressed the average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), and (3) the interaction between (1) and (2).

Table S3.5 presents the results for each condition. Sliders led to significantly greater error and was thus less accurate for both the right-skewed and the symmetric distribution, $b = 0.13$, $SE = 0.05$, $p = .004$. Neither the main effect of distribution shape ($b = -0.08$, $SE = 0.05$, $p = .098$) nor the interaction effect was significant ($b = 0.02$, $SE = 0.09$, $p = .800$).

Table S3.5. Study 5 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Right-skewed	**1.59**	0.78	1.73	0.95	$t(749) = -2.30, p = .022$
Symmetric	1.68	0.85	1.80	1.04	$t(783) = -1.80, p = .072$

Note. Within each row, **boldface** indicates that this method generates significantly smaller AAE and is significantly more accurate.

In sum, Study 5 shows that Distribution Builder was more accurate in eliciting implicitly memorized information, both when the true distribution had a right-skewed shape and a symmetric shape.

Study 6

Our dependent measure was average absolute error, calculated the same way as in previous studies. To test whether the accuracy depends on the elicitation method, the distribution shape, and the perception of numbers, we regressed the average absolute error on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), (3) the random number condition (contrast-coded), (4) the interaction between (1) and (2), (5) the interaction between (1) and (3), (6) the interaction between (2) and (3), and (7) the three-way interaction between (1), (2) and (3). We included fixed effects for each task and clustered standard errors by participant.

Table S3.6 displays the results for each condition. Consistent with Study 5, we found that Distribution Builder generated significantly smaller error than Sliders regardless of the distribution shape and regardless of whether the numbers were random or meaningful, b = -0.26, SE = 0.05, p < .001. The main effect of distribution shape was significant such that elicited distributions had significantly smaller error for right-skewed than left-skewed distribution, b = -0.15, SE = 0.04, p < .001. Whether the stimuli were presented as meaningful or random numbers did not influence the result, b = 0.04, SE = 0.05, p = .389. None of the interactions was significant (ps ≥ .264).

Table S3.6. Study 6 Results: Average Absoluate Error of Distribution Builder Versus Sliders Condition

	Distribution Builder		Sliders		
	M	*SD*	*M*	*SD*	Paired t-test
Right-skewed / Meaningful Numbers	**2.42**	1.16	2.61	1.27	$t(953) = -2.38$, $p = .017$
Left-skewed / Meaningful Numbers	**2.55**	1.24	2.86	1.45	$t(954) = -3.57$, $p < .001$
Right-skewed / Random Numbers	**2.47**	1.28	2.73	1.36	$t(942) = -3.07$, $p = .002$
Left-skewed / Random Numbers	**2.57**	1.33	2.85	1.57	$t(944) = -2.93$, $p = .004$

Note. Within each row, **boldface** indicates that this method generates significantly smaller AAE and is significantly more accurate.

Taken together, Study 6 suggests that for both right-skewed and left-skewed distributions and for both random and meaningful stimuli in an explicit memorization task, Distribution Builder consistently produced smaller errors than Sliders.

Study 7

To test whether the accuracy of the subjective distribution depends on the elicitation method and the skewness of the answer, we regressed the accuracy measure on (1) the Sliders condition (contrast-coded), (2) the right-skewed condition (contrast-coded), and (3) the interaction between (1) and (2).

Table S3.7 presents the results for each task. Across almost all tasks, participants in the Distribution Builder condition allocated significantly more mass to the correct bin and were thus more accurate both when the response pattern was right-skewed and symmetric, $b = 1.85$, $SE = 0.56$, $p = .001$. The only exception was in the age estimation task, where Sliders were directionally more accurate than Distribution Builder in the right-skewed condition, leading to a significant interaction between method and skewness, $b = 4.66$, $SE = 2.07$, $p = .024$.

In addition, a right-skewed response pattern led to higher accuracy in elicited subjective distributions, $b = 4.73$, $SE = 0.54$, $p < .001$. With all tasks combined, the interaction between method and response shape was not significant, $b = 1.38$, $SE = 1.09$, $p = .205$.

Table S3.7. Study 7 Results: Percentage Allocated to the Correct Bin of Distribution Builder Versus Slider Condition

	Distribution Builder		Slider		
	M	*SD*	*M*	*SD*	Paired t-test
Weight Distribution					
Right-skewed	28.63	24.81	27.25	22.81	$t(785) = 0.81, p = .416$
Symmetric	33.08	22.71	31.54	23.01	$t(774) = 0.93, p = .351$
Age Distribution					
Right-skewed	37.90	19.96	39.32	21.82	$t(777) = -0.95, p = .344$
Symmetric	**33.21**	19.39	29.96	20.41	$t(782) = 2.27, p = .023$
Temperature Distribution					
Right-skewed	**35.80**	18.55	33.07	18.48	$t(779) = 2.06, p = .040$
Symmetric	**28.61**	18.98	25.44	19.05	$t(780) = 2.34, p = .020$
Walking Time Distribution					
Right-skewed	29.98	22.01	27.89	20.82	$t(784) = 1.37, p = .172$
Symmetric	21.25	19.58	18.89	18.47	$t(775) = 1.73, p = .084$
All Tasks Combined					
Right-skewed	33.05	21.82	31.87	21.56	$t(3,131) = 1.52, p = .129$
Symmetric	**29.02**	20.76	26.49	20.89	$t(3,117) = 3.38, p < .001$

Note. Within each row, **boldface** indicates that participants using this method allocate significantly more mass to the correct bin and is significantly more accurate.

Consistent with results in Studies 5 and 6, the current study suggests that the advantage of Distribution Builder over Sliders persists even in the domain of subjective probability judgments, regardless of whether the correct answer falls at the lower end or the middle of the range.

Supplement 4: Study S1

In Study S1, we investigated whether people generate more accurate belief distributions when presenting the Sliders interface in a vertical orientation. In addition to the Distribution Builder and the (standard) Horizontal Sliders condition, we added a Vertical Sliders condition. We used the same study procedure as Study 8, in which participants constructed belief distributions for specific prediction questions and indicated their confidence using two other measures.

Method

Participants. We pre-registered to recruit 3,000 U.S. participants from MTurk. Participants (n = 20) who failed an attention check question at the beginning of the survey were automatically excluded. As pre-registered, we excluded duplicate responses from the same IP addresses (59 exclusions) or the same MTurk IDs (11 exclusions) and participants who misreported their MTurk IDs (33 exclusions). The exclusions left us with a final sample of 2,891, which averaged 41.0 years of age[1]; 1,614 reported their gender as female (55.9%), 1,243 reported as male (43.0%), 18 reported a gender not listed (0.6%), and 13 preferred not to disclose (0.5%).

Procedure. We used the same study procedure as Study 8. All participants first answered four binary preference and behavior questions themselves and were then randomly assigned to predict for two of these questions the percentage of all respondents who would choose a particular option by providing their entire belief distribution. We randomly assigned participants to use the Distribution Builder, the Horizontal Sliders, or the Vertical Sliders for both questions. The interfaces for the Distribution Builder and the Horizontal Sliders were the same as in previous

[1] Participants reported their age in an open-ended text box. When calculating the average age in each study, we excluded any response that was lower than 18 or higher than 100.

studies. The only difference between the Vertical Sliders and the Horizontal Sliders was the orientation (see Figure S1). Participants saw the same labeled bins with a vertical slider scale above each of them and adjusted the sliders from bottom to top to indicate their confidence that the true answer fell within each bin.

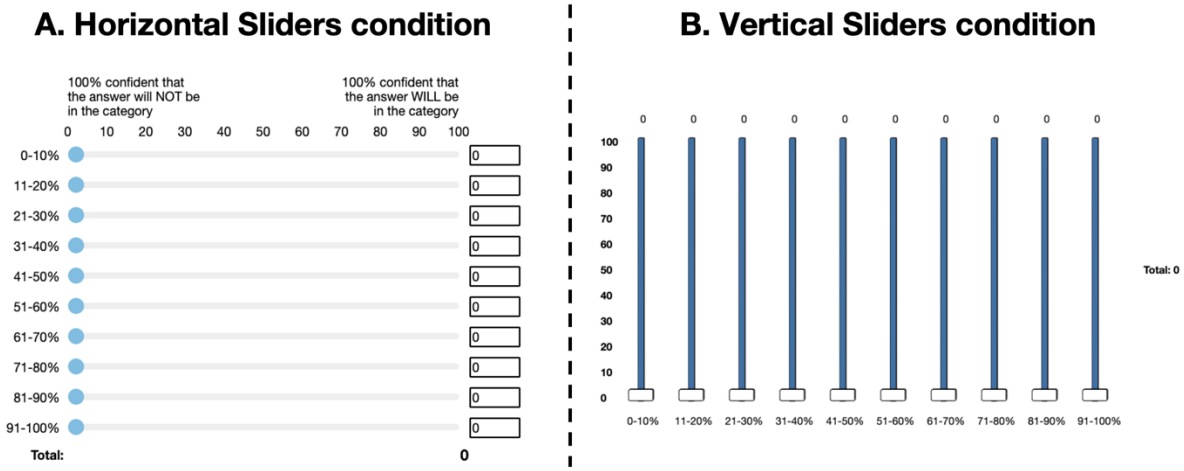

Figure S1. Distribution elicitation page presented in Study S1 (Panel A: (Standard) Horizontal Sliders condition; Panel B: Vertical Sliders condition).

As in Study 8, after providing their belief distributions, all participants predicted the option that more survey respondents would choose and reported how confident they were in their binary prediction. For each question, they also received a $0.20 bonus and decided how much of it to wager on this prediction. The amount they wagered would double if their prediction was correct.

Results and Discussion

As in Study 8, we calculated participants' confidence implied by the belief distribution as the percentage of mass they allocated to the half of the range corresponding to their prediction. We

then derived the absolute difference between participants' response to the confidence question and their confidence implied by the belief distribution. For this measure, we pre-registered the following two regressions: First, we used OLS to regress the measure on (1) the Horizontal Sliders condition (dummy-coded) and (2) the Vertical Sliders condition (dummy-coded). Second, we used OLS to regress the measure on (1) the Distribution Builder condition (dummy-coded) and (2) the Vertical Sliders condition (dummy-coded). Both regressions included fixed effects for the prediction question and clustered standard error by participants.

Compared to Distribution Builder ($M = 19.49$, $SD = 16.52$), both Horizontal Sliders ($M = 22.20$, $SD = 18.81$) and Vertical Sliders ($M = 23.35$, $SD = 19.37$) yielded significantly greater deviation between participants' self-reported confidence and their confidence implied by the belief distribution ($b = 2.71$, $SE = 0.63$, $p < .001$ for Horizontal Sliders, $b = 3.85$, $SE = .64$, $p < .001$ for Vertical Sliders). Rotating the orientation of the Sliders interface to be vertical instead of horizontal did not reduce (and directionally increased) the difference between beliefs implied by elicited distributions and directly stated beliefs, $b = 1.14$, $SE = 0.68$, $p = .093$.

We also pre-registered to analyze how well the confidence implied by elicited distributions is predicted by participants' wager amounts. We conducted two sets of regressions: First, we used OLS to regress this measure on (1) the Horizontal Sliders condition (dummy-coded); (2) the Vertical Sliders condition (dummy-coded); (3) the wager amount (mean-centered); (4) the interaction between (1) and (3); and (5) the interaction between (2) and (3). We included fixed effects for the prediction question and clustered standard errors by participant. We are primarily interested in the two interaction terms. The first interaction term examined whether the correlation between wager amount and confidence was different in the Horizontal Sliders condition than in the Distribution Builder condition. The second interaction term examined whether this correlation

was different in the Vertical Sliders condition than in the Distribution Builder condition. Second, we used OLS to regress the same measure on (1) the Distribution Builder condition (dummy-coded); (2) the Vertical Sliders condition (dummy-coded); (3) the wager amount (mean-centered); (4) the interaction between (1) and (3); and (5) the interaction between (2) and (3). We included fixed effects for the prediction question and clustered standard errors by participant. Similarly, the first interaction term examined whether the correlation between wager amount and confidence was different in the Distribution Builder condition than in the Horizontal Sliders condition. The second interaction term examined whether this correlation was different in the Vertical Sliders condition than in the Horizontal Sliders condition.

Table S4.1 presents the regression results. As previously mentioned, the interaction terms reflect whether the correlation between wager amount and confidence implied by belief distributions differ across conditions. This correlation was significantly different between the Distribution Builder condition ($r = .14$, $p < .001$) and the Vertical Sliders condition ($r = .05$, $p = .019$), $p = .015$. But the correlation in the Horizontal Sliders condition ($r = .11$, $p < .001$) did not significantly differ from the other two conditions ($ps \geq .105$).

Table S4.1. Study S1 Results

Variables	(1) OLS	(2) OLS
Horizontal Sliders (dummy-coded)	-5.35***	
	(0.918)	
Vertical Sliders (dummy-coded)	-4.17***	1.18
	(0.928)	(0.984)
Distribution Builder (dummy-coded)		5.35***
		(0.918)
Wager amount (mean-centered)	0.52***	0.42***
	(0.087)	(0.092)
Horizontal Sliders * Wager amount	-0.10	
	(0.127)	
Vertical Sliders * Wager amount	-0.31*	-0.21
	(0.129)	(0.132)
Distribution Builder * Wager amount		0.10
		(0.127)
Fixed effects for prediction questions	Yes	Yes
Constant	70.2***	64.86***
	(0.848)	(0.905)
Observations	5,782	5,782

Note. Robust standard errors in parentheses. The dependent measure in both regressions is the confidence implied by the belief distribution, which equals to the amount of mass participants allocated to the half of the interface corresponding to their binary prediction.

As in Study 8, we also derived as an exploratory measure the proportion of participants who did not allocate more mass to the half corresponding to their prediction. Consistent with other measures, this proportion was significantly higher in both the Horizontal Sliders condition (M = 26.92%, SD = 0.44) and the Vertical Sliders condition (M = 28.77%, SD = 0.45) than in the Distribution Builder condition (M = 19.58%, SD = 0.40), ps < .001 for both comparisons, suggesting that participants in both Sliders conditions produced more erroneous distributions.

Taken together, these results suggest that rotating the Sliders orientation to vertical did not improve Sliders' performance. In fact, the Vertical Sliders in this study did directionally worse than the Horizontal Sliders on both measures.

We suggest in the paper that less accurate results elicited by the Sliders may be in part due to starting bin when participants construct the distributions. In other studies (see Table 3 in the paper), participants were more likely to start from the first bin and thus put more mass in that first bin when using the Sliders interface. In the current study, we do not have data on the order in which participants built distributions in the Vertical Sliders condition. Nevertheless, the data we do have

regarding how participants allocated mass across the range are consistent with the results we present in the paper. Table S4.2 suggests that compared to those using Distribution Builder, participants using Sliders tend to allocate significantly more mass into the first bin and the first half of the distribution, regardless of whether the Sliders interface was orientated vertically or horizontally.

Table S4.2. Allocation Measures for Study S1

Distribution Builder	Horizontal Sliders	Vertical Sliders	Effect of the Horizontal Sliders condition (vs. Distribution Builder condition)	Effect of the Vertical Sliders condition (vs. Distribution Builder condition)
% of Mass Allocated to the First Bin				
3.06%	5.58%	4.26%	$b = 2.51$, $SE = 0.40$, $p < .001$	$b = 1.21$, $SE = 0.37$, $p = .001$
% of Mass Allocated to the First Half of the Distribution				
48.48%	54.30%	51.40%	$b = 5.74$, $SE = 1.06$, $p < .001$	$b = 3.04$, $SE = 1.10$, $p = .006$

Supplement 5: Study S2

One possible reason why Distribution Builder users generate more accurate distributions is that Distribution Builder requires more effortful (one-ball-at-a-time) inputs than Sliders, which may prompt its users to be more deliberate about their responses. If this is true, if a lack of deliberation leads Sliders users to be less accurate, then we should be able to make them more accurate by incentivizing them to revise their answers in an attempt to be more accurate. We tested this in Study S2.

In this study, we used an explicit memorization procedure, similar to the one used in Studies 1-4, 6, and 10, and randomly assigned participants to use either Distribution Builder or Sliders. After all participants submitted their (unincentivized) distributions, we offered them an incentive to revise their distributions to make them more accurate. If the accuracy difference between Sliders and Distribution Builder is caused by reduced deliberation among Sliders users, then incentivizing participants to deliberate more when constructing their distributions should affect Sliders users more than Distribution Builder users, thereby reducing the accuracy gap between these two methods.

Method

Participants. We pre-registered to recruit 1,000 U.S. participants from Prolific. Participants who failed the comprehension check twice (n = 2) were automatically excluded from the survey. As pre-registered, we excluded duplicate responses from the same IP addresses (18 exclusions) or the same Prolific IDs (0 exclusions) and participants who misreported Prolific IDs (3 exclusions). We also excluded participants with an average absolute error of 0 at either stage (3 exclusions). These exclusions left us with a final sample of 976, which averaged 30.1 years of age; 764 reported

their gender as female (78.3%), 190 reported as male (19.5%), 21 reported a gender not listed (2.2%), and 1 preferred not to disclose (0.1%).

Procedure. We randomly assigned participants to one of the two conditions in a 2 (method: Distribution Builder vs. Sliders) × 2 (response: unincentivized first response vs. incentivized second response) mixed design with method as the between-subjects factor and response as the within-subjects factor.

The first part of the study is very similar to other studies in the paper that used an explicit memorization task. All participants first viewed 50 movie ratings (min. = 1, max. = 50), randomly selected from 200 numbers generated from a normal distribution (M = 25, SD = 10; truncated at min. = 1 and max. = 50). The 50 values were presented one at a time, at a fast pace. After viewing the 50 movie ratings, participants were randomly assigned to reproduce the distribution of the 50 values using either Distribution Builder or Sliders. All participants provided their first response without any incentive for accuracy.

After submitting their first response, participants learned that they would have the opportunity to revise their previous answers and could earn a bonus payment of up to $1, depending on the accuracy of their final submission. On the next page, participants saw their previous responses pre-populated to the elicitation interface and were allowed to make any changes before submitting their second response. Figure S2 presents an example in the Distribution Builder condition. Finally, participants reported their age and gender.

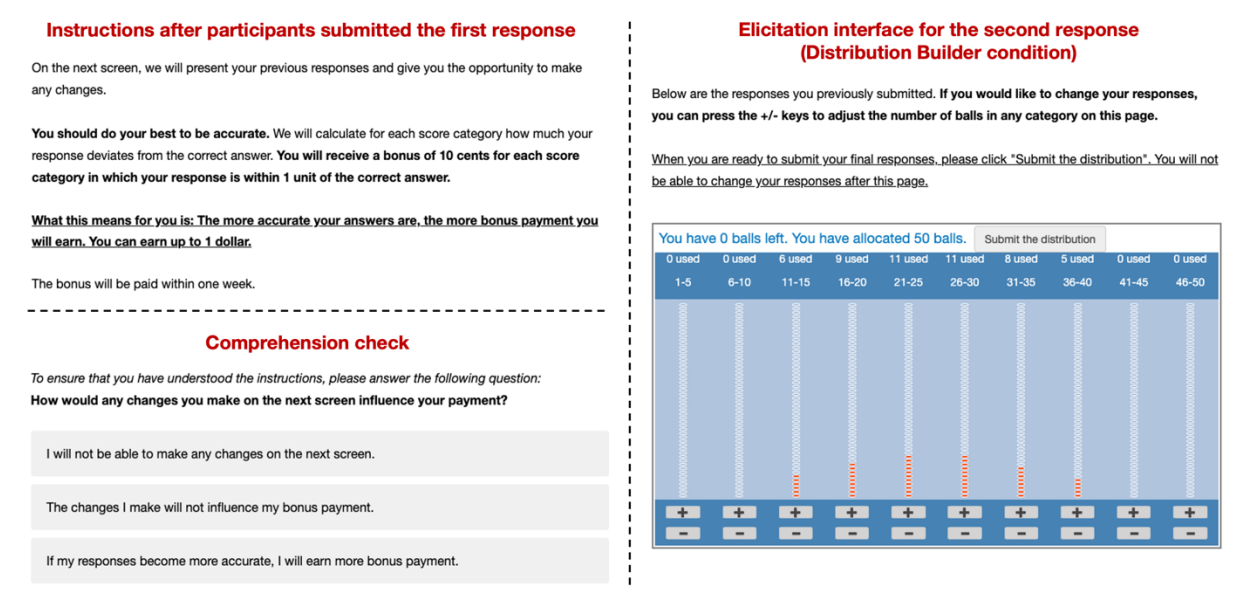

Figure S2. Instructions for the second, incentivized response in the Distribution Builder condition in Study S2. Dashed lines indicate page breaks.

Results and Discussion

As in other memory task studies, we calculated, for each bin, the absolute difference between the frequency participants indicated and the true frequency presented. We then averaged the absolute difference across all bins to obtain a measure of average absolute error. Each participant contributed two rows to the dataset, one for each response. We pre-registered to use OLS to regress the average absolute error on (1) the Sliders condition (contrast-coded), (2) incentivized second response condition (contrast-coded), (3) their interaction. We clustered standard errors by participant.

Replicating previous results, we found that Slider responses (M = 2.36, SD = 1.18) had significantly greater error than Distribution Builder responses (M = 2.22, SD = 1.02), b = 0.14, SE = 0.07, p = .038. Reassuringly, we also found a main effect of revision, such that participants' second responses had significantly less error than their first responses, b = -0.10, SE = 0.01, p

< .001. However, we found no hint of an interaction between the two factors, $b < 0.001$, $SE = 0.02$, $p = .974$. That is, while the incentive to revise responses reduced error for both Sliders ($b = -0.10$, $SE = 0.02$, $p < .001$) and Distribution Builder ($b = -0.10$, $SE = 0.02$, $p < .001$), it helped both methods to a similar degree so that Sliders responses remained less accurate than Distribution Builder responses after the revision. The results hold when we restrict our analyses to the subset of participants who correctly answered the comprehension question about the incentive on the first try ($n = 970$) and when we only analyze participants who did revise their responses ($n = 421$).

We also derived and analyzed a few secondary variables, and those yielded consistent results. The two conditions did not differ in (1) the amount of total adjustment, calculated as the absolute deviation between the first and second responses (Distribution Builder: $M = 2.60$, $SD = 4.89$; Sliders: $M = 3.05$, $SD = 6.08$; $b = 0.45$, $SE = 0.35$, $p = .204$), or (2) the percentage of participants who improved accuracy, or, in other words, had smaller average absolute error in the second response (Distribution Builder: $M = 23.06\%$, $SD = 0.42$, Sliders: $M = 24.90\%$, $SD = 0.43$; $b = 0.02$, $SE = 0.03$, $p = .502$). These results are consistent with our primary results that the opportunity of revision improved both methods in a similar way and did not help Sliders users more than Distribution Builder users.

Nevertheless, it is worth noting that while the first response by Sliders users ($M = 2.41$, $SD = 1.21$) had greater error than the first response by Distribution Builder users ($M = 2.27$, $SD = 1.06$), $t(974) = 1.98$, $p = .048$, the *second* response by Sliders users ($M = 2.31$, $SD = 1.14$) was no less accurate than the *first* response by Distribution Builder users, $t(974) = 0.67$, $p = .503$. That is, after revision, Sliders users' responses became just as accurate as the initial responses in the Distribution Builder condition. This suggests that if, for any reason, researchers decide to use Sliders to elicit belief distributions, incentivizing participants to revise their responses might increase response

accuracy to the extent that they become as accurate as unincentivized Distribution Builder responses. There is one practical caveat to point out. Researchers can't incentivize participants to generate accurate distributions when they don't know what an accurate distribution is. So this is only applicable in contexts in which accurate distributions are known, such as the memory context studied here.

Supplement 6: Study S3

Study S3 aimed to investigate whether instructing participants to start from the bin they think is most likely will reduce the bias in using Sliders. Before constructing the distribution, participants first chose the bin that they thought should contain the most mass, and we instructed them to start from that bin when filling out the distribution.

Method

Participants. We pre-registered to recruit 1,200 U.S. participants from the CloudResearch-approved pool on MTurk. As pre-registered, participants who failed the comprehension check twice at the beginning of the survey (n = 28) were automatically excluded from the study. For duplicate responses from the same MTurk IDs or the same IP addresses, we included the first responses only (17 exclusions). We excluded 2 participants with average absolute error of zero. The exclusions left us with a final sample of 1,183, which averaged 42.8 years of age; 619 reported their gender as female (52.3%), 550 as male (46.5%), 9 reported a gender not listed (0.8%), and 5 preferred not to disclose (0.4%).

Procedure. Participants were randomly assigned to one of the four conditions in a 2 (method: Distribution Builder vs. Sliders) × 2 (order instructions: yes vs. no) between-subjects design. We used a similar memorization procedure as in Studies 1-4, 6, 10, and S2. Participants read that they would observe the resting heart rate of 50 randomly selected adults, with rapid succession. All participants saw 50 numbers from a left-skewed distribution, which are the same as in the left-skewed condition in Study 6. We only included stimuli from a left-skewed distribution because

our past studies suggested that this is where the bias in using Sliders is more likely to manifest. We manipulated the elicitation method the same way as in all other studies.

We additionally manipulated whether participants were instructed to start from the modal bin or not. In the Control (i.e., no order instructions) condition, the procedure was exactly the same as in other studies: Participants proceeded to the elicitation interface after observing the 50 numbers. In the Order instructions condition, participants were first shown a page with all possible bins and chose the one they thought contained the most numbers they saw. Then, on a separate page, they saw a screenshot of the Distribution Builder or Sliders interface with the chosen bin circled and read that they should start with that bin when constructing the distribution on the following page. To ensure that participants read and understood this instruction, we asked them to type the following sentence into a text box before proceeding to the elicitation page: "I understand that, on the next page, I will start with the '[chosen]' category when allocating the 50 balls" in the Distribution Builder condition or "I understand that, on the next page, I will start with the '[chosen]' category when moving the sliders" in the Sliders condition.

Results and Discussion

As in all studies with memory task in the paper, our dependent measure is average absolute error. We used OLS to regress the average absolute error on (1) the Sliders condition (contrast-coded), (2) No Order Instructions condition (contrast-coded), and (3) the interaction between (1) and (2).

First, as a manipulation check, we looked at which bin participants started from when building the distributions. In the Order Instructions condition, most participants (85.71% in the Distribution Builder condition, 79.58% in the Sliders condition) followed the instructions and started from the

bin they chose to contain the most mass. Replicating the order results in Table 3 in the paper, when there were no order instructions, significantly more Sliders users started from the first bin (M = 57.05%, SD = 0.50) than Distribution Builder users (M = 42.32%, SD = 0.49), b = 0.15, SE = 0.04, $p < .001$. This discrepancy persists even with the order instructions (Distribution Builder: M = 6.12%, SD = 0.24; Sliders: M = 11.97%, SD = 0.33; b = 0.06, SE = 0.02, p = .014).

Without the order instructions, Sliders (M = 2.95, SD = 1.52) led to significantly greater error than Distribution Builder (M = 2.71, SD = 1.28), b = 0.24, SE = 0.11, p = .036. To our surprise, the order intervention significantly increased error, and by a lot: b = 0.63, SE = 0.10, $p < .001$. Within the order instructions condition, Sliders (M = 3.56, SD = 1.93) still led to directionally greater error than Distribution Builder (M = 3.36, SD = 1.86), b = 0.20, SE = 0.16, p = .208. Overall, we did not find a significant interaction between elicitation method and the order intervention, b = 0.04, SE = 0.19, p = .830, and a main effect of method suggests that Sliders led to significantly greater error overall, b = 0.22, SE = 0.10, p = .023.

Thus, at least in the context of memorization tasks, the intervention of changing the order in which people construct distributions may backfire and reduce response accuracy for all participants. Exploratory analyses reveal one possible explanation: Even though participants indicated the correct modal bin, forcing them to start from that bin may have led them to allocate too much mass into that bin in the first place. We derived the amount of mass participants allocated to the modal bin in their distribution and found that those in the order instructions condition (M = 34.30%, SD = 0.15) allocated significantly more mass to the modal bin than those in the control condition (M = 26.39%, SD = 0.07), $t(1,181) = 12.04$, $p < .001$. This may have in turn increased their response errors, as in the "ground truth" distribution presented to participants, the modal bin only had 22% of the mass.