Memory and Representativeness

Bordalo, Coffman, Gennaioli, Schwerter, Shleifer

Supplementary Materials

A Derivation of the Predictions

To derive the predictions, we start from Expression (3) for the average assessed probability of hypothesis $h =$ orange given data $d =$ numbers, and obtain:

$$\widetilde{P}(h|d) = \frac{e^{S(d,h)}}{e^{S(d,h)} + e^{S(d,-h)}} = \frac{1}{1 + e^{S(d,-h)-S(d,h)}}$$

Thus, $\widetilde{P}(h|d)$ is a monotonically increasing function of

$$S(d,h) - S(d,-h) = \alpha\left[f(P(d,h)) - f(P(d,-h))\right] - \gamma\left[f(P(-d,h)) - f(P(-d,-h))\right]$$

It follows that:

$$\frac{d\widetilde{P}(h|d)}{dP(h|d)} \propto \alpha\left[f'(P(h|d)P(d)) + f'((1-P(h|d))P(d))\right]P(d)$$

which is positive if and only if $\alpha > 0$, because $f()$ is an increasing function. Similarly, we find:

$$\frac{d\widetilde{P}(h|d)}{dP(h|-d)} \propto -\gamma\left[f'(P(h|-d)P(-d)) + f'((1-P(h|-d))P(-d))\right]P(-d)$$

which is negative if and only if $\gamma > 0$.

Predictions 1 and 3 then follow from the latter comparative static. For Prediction 1, $P(blue|words) > P(blue|shapes) = 0$ so that

$$\widetilde{P}(blue|numbers)_{d=words} < \widetilde{P}(blue|numbers)_{d=shapes}$$

For Prediction 3, $P(blue|words)_k$ decreases in the number k of orange words, so that $\widetilde{P}(blue|numbers)_k$ increases in that number.

Prediction 2 reflects the interpretation of Equation (3) as the probability with which hypothesis h comes to mind when cued with data d.

Finally, Prediction 4 reflects the role of the cue. Formally, denote the hypothesis space by $H \times G$, so that a complete hypothesis is a pair (h, g). In our setting, given data d or $-d$, each $h \in H$ only occurs with a single $g \in G$ and vice versa. When the cue is H (e.g. color), denote the probability of hypothesis h with data $-d$ as $P(h(g)|-d)$.

When the cue is G (e.g. size), denote the probability of hypothesis g with data $-d$ as $P(g(h)|-d)$. Prediction 4 then follows from the fact that $P(blue(large)|words) > P(small(orange)|words) = 0$.

B Methods, Procedures, and Further Results

B.1 Study 1

Procedural details and data collection

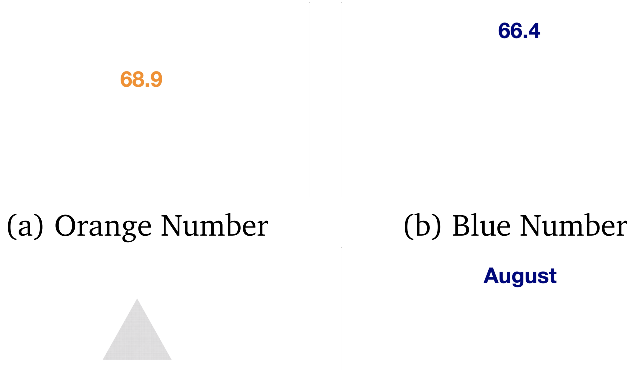
In this section, we provide further details about Study 1. We conducted two waves of the *blue* and *gray* treatments. These experiments were conducted in May of 2018 in the laboratories of the University of Cologne (N=427) and at Bocconi University (N=363). The entire lab experiment took 10 minutes. Subjects received a show-up fee of €4.00. In case they were randomly selected to receive additional payments based on our main treatments on memory and representativeness, subjects received €0.50 for each correct answer to the questions on the 50 images (see the full list of questions below). A distraction task on emotional expressions was used in the wave conducted at the University of Cologne and took 90 seconds on average. A distraction task on raven matrices was used in the wave conducted at Bocconi University and took 170 seconds on average. We find no treatment differences between the two waves and hence present in the main text the results of the treatments with distraction by pooling both waves. In the following we also show the non-pooled results.

Images

Screenshots for each type of image used in Study 1 are displayed in Figure B.1. Figures B.2 and B.3 list the target images (orange and blue numbers) and decoy images (blue words or gray shapes, depending on treatment).

Questions

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely color of the chosen image? Blue or Orange.
- Q2: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the probability that this number is orange?



(a) Orange Number (b) Blue Number

(c) Gray Shape (d) Blue Word

Figure B.1: Examples of image screenshots

Orange Numbers:

67.8	68.9	72.6	73.0	54.0
57.1	65.2	52.4	56.2	61.3,

Blue Numbers:

66.4,	51.3	58.5	71.5	69.5
56.7,	62.7	50.8	70.8	59.6,
74.3,	53.9	63.6	60.0	64.1

Figure B.2: Full set of target images shown to participants

- Q3: How many orange numbers were shown to you?
- Q4: How many blue numbers were shown to you?
- *Blue* treatment only:

AddQ1: How many blue words were shown to you?

AddQ2: How many orange numbers were shown to you?

- *Gray* treatment only:

AddQ1: How many gray shapes were shown to you?

- Q5: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the probability that this number is blue?

Blue Words:

August	May	June	July	Millenium
Hour	Month	November	Year	March
Tuesday	December	Second	Saturday	Decade
January	Week	Minute	October	Monday
Sunday	Century	Day	Wednesday	April

Gray Shapes:

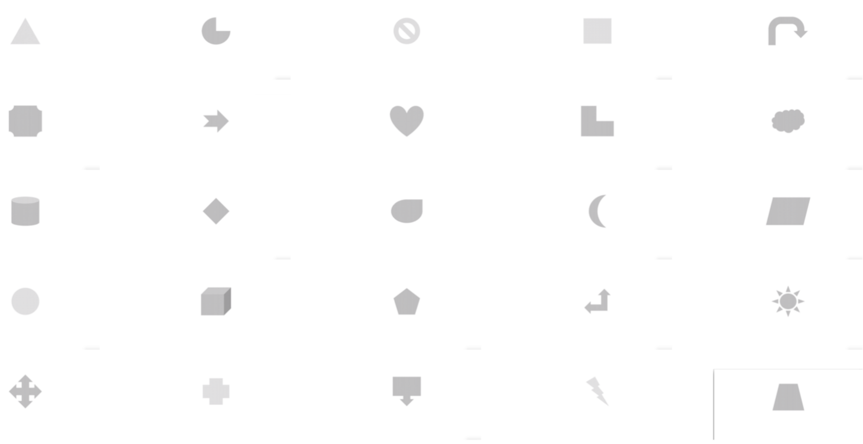


Figure B.3: Decoy images shown to participants depending on treatment

- The question was used only in the second wave of Study 1. Because of a computer error, we have responses to this question only for roughly half of the 2nd wave's sample.

Distraction tasks After participants were exposed to the 50 images and before they responded to the questions, we employed one of two distraction tasks to wash out participants working memory, an emotion expression task (adapted from a quiz created by The Greater Good Science Center at UC Berkeley (https://greatergood.berkeley.edu/quizzes/take_quiz/ei_quiz), and a raven matrices task (see Figure B.4 for an example). The goal of the distraction task is to fully engage participants' working memory in a task orthogonal to the key task at hand (recall of the images). By presenting this distraction task, we likely impose a heavy extraneous cognitive load (see Paas, Renkl, and Sweller 2003), swamping participants' working memories with new information. If our results persist after the distraction, this suggests a role for recall rather than working memory.

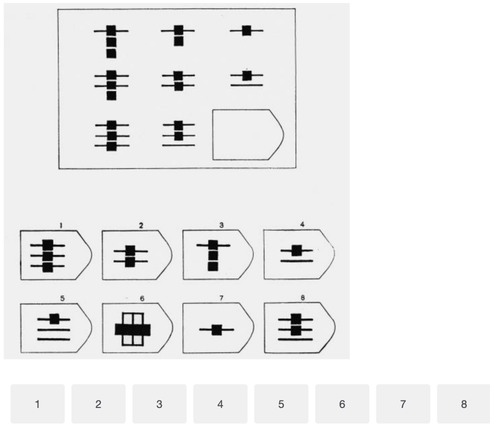


Figure B.4: One of the tasks of the raven matrices used to distract participants in wave 2 of Study 1.

Further results

Our main results of Study 1 reported in Table 3 also hold when (i) providing different statistical tests (see Table B.1) and (ii) looking at our two waves separately (see Table B.2 for wave 1 and Table B.3 for wave 2).

Table B.1: Regression estimates of treatment effects in Study 1

	Logit: Y=1 if "orange is more likely"	Logit: Y= 1 if more orange numbers recalled	OLS Y= Orange numbers recalled	OLS Y= Blue numbers recalled	OLS Y= Ratio of orange to total numbers recalled	OLS Y= Probability that a randomly-drawn number is orange	OLS Y= Probability that a randomly-drawn number is blue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 if *blue*	.8970***	.7741***	.3807	− 1.200*	.0412***	.0613***	−.0296
	(.1498)	(.1564)	(.6011)	(.6828)	(.0116)	(.0134)	(.0332)
Wave dummy	yes	yes	yes	yes	yes	yes	no
Constant	−.8008***	−1.070***	13.57***	16.29***	.4575***	.4466***	.5498***
	(.1270)	(.1341)	(.4990)	(.5668)	(.0096)	(.0111)	(.0253)
Observations	790	790	790	790	790	790	117
Adj./Ps. R^2	0.04	0.03	0.00	0.00	0.01	0.03	0.00

Table B.2: Regression estimates of treatment effects in Study 1b for wave 1

	OLS: Y=1 if "orange is more likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg: Y= Ratio of orange to total numbers recalled	0.5-Q-Reg: Y= Probability that a randomly-drawn number is orange
	(1)	(2)	(3)	(4)	(5)	(6)
1 if *blue*	.2445***	.1615***	0	0	.0455**	.05*
	(.0465)	(.0452)	(.9386)	(.7426)	(.0217)	(.0277)
Constant	.3049***	.2601***	12***	15***	.4545***	.45***
	(.0322)	(.0313)	(.6488)	(.5132)	(.0150)	(.0192)
Observations	427	427	427	427	427	427
(Ps.) R^2	0.05	0.03	0.00	0.00	0.01	0.01

Table B.3: Regression estimates of treatment effects in Study 1b for wave 2

	OLS: Y=1 if "orange is more likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg: Y= Ratio of orange to total numbers recalled	0.5-Q-Reg: Y= Probability that a randomly-drawn number is orange	0.5-Q-Reg: Y= Probability that a randomly-drawn number is blue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1 if *blue*	.1954***	.1728***	0	−2**	.0806***	.1***	−.1***
	(.0504)	(.0480)	(.6960)	(.7859)	(.0192)	(.0251)	(.0373)
Constant	.2840***	.2189***	10***	14***	.4194***	.4***	.6***
	(.0368)	(.0351)	(.5088)	(.5745)	(.0141)	(.0184)	(.0285)
Observations	363	363	363	363	363	363	117
(Ps.) R^2	0.04	0.03	0.00	0.01	0.02	0.03	0.02

B.1.1 Study 1b

We also ran Study 1 in the lab and on MTurk without distraction between the image presentation and questions and without eliciting Q2, called Study 1b. Here is a link to an online version of the two treatments of Study 1b:

https://unikoelnwiso.eu.qualtrics.com/jfe/form/SV_6PaEfC4u67hWp1P

Procedural details and data collection

We conducted Study 1b in three waves. Wave 1 was conducted in February of 2018 with MTurk and a sample of 337 participants. Our *blue* and *gray* treatments were accompanied by an unrelated intertemporal choice. The entire experiment lasted for around 10 minutes. Participants received a $1.00 show-up fee. A computer-based coin toss determined randomly whether subjects would receive additional payments based on the *blue* and *gray* treatments or on the unrelated intertemporal choice. In case subjects received additional payments based on the former, one of all participants was randomly chosen to receive $20.00 for each correct answer to Q1, Q3, Q4 and AddQ1-2, while all remaining participants received $0.20 for each correct answer.

We then successfully replicated our findings in two more waves, one in the laboratory of the University of Cologne in March of 2018 (N=483) and another with MTurk in March 2018 (N=193). In our first replication, we tested whether our results are robust to moving from MTurk to the laboratory. The laboratory offered us more control on the image display, because we could ensure equally stable internet connection and computing power for each participant. Like in the first wave, the *blue* and *gray* treatments were accompanied by an unrelated intertemporal choice. The entire lab experiment also took 10 minutes. Subjects received a show-up fee of €4.00. One participant per experimental session (consisting of 26 to 32 participants) was randomly selected to receive additional payments based on the intertemporal choice task. All remaining participants received € 0.50 for each correct answer to Q1, Q3, Q4 and AddQ1-2 of the *blue* and *gray* treatments.

In our final replication, we tested whether our results are robust to conducting the *blue* and *gray* treatments without being accompanied by unrelated intertemporal choice tests. The experiment of the third wave took below 7 minutes. Subjects received a $0.50 show-up fee and $0.20 for each correct answer to Q1, Q3, Q4, and AddQ1-2.

Questions

In order of exposure to participants:

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely color of the chosen image? Blue or Orange.
- Q3: How many orange numbers were shown to you?
- Q4: How many blue numbers were shown to you?
- *Blue* treatment only:

AddQ1: How many blue words were shown to you?

AddQ2: How many orange numbers were shown to you?

- *Gray* treatment only:

AddQ1: How many gray shapes were shown to you?

Main Results of Study 1b

Further results

The results are summarized in Figure B.5 and Table B.4 and show strong support for Prediction 1. Column (1) in Table B.4 reports an OLS regression of a response dummy (1 if "orange is likely") on a treatment dummy (1 if $blue$), that amounts to comparing the average share of participants who said orange is likely in the $gray$ treatment versus the $blue$ treatment. As shown in Figure B.5, that share increases 20.8pp from the $gray$ treatment to the $blue$ treatment (35.3% to 56%, significant at 1% level).

Column (2) reports the results of an OLS regression of a response dummy that takes value 1 if the participant reported more orange numbers in Q3 than blue numbers in Q4, which is an alternative measure of each participants' belief about the likely color. Consistent with Column (1), there is a 13.8pp increase in the share of participants who

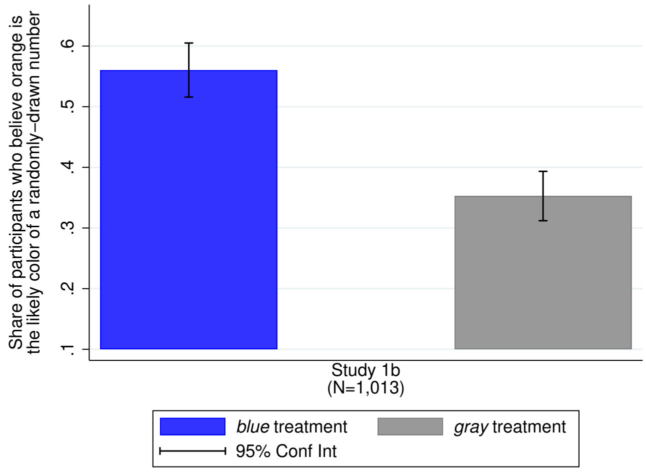


Figure B.5: Share of participants who believe that the likely color of a randomly-drawn number is orange for the *blue* and *gray* treatments of Study 1.

Table B.4: Regression estimates of treatment effects in Study 1b

	OLS: Y=1 if "orange is likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg: Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.2060***	.1379***	0	− 2***	.0556***
	(.0307)	(.0302)	(.4187)	(.6427)	(.0124)
MTurk dummy	yes	yes	yes	yes	yes
Wave dummies	yes	yes	yes	yes	yes
Constant	.3305***	.3043***	10***	14***	.4444***
	(.0302)	(.0259)	(.3598)	(.5524)	(.0107)
Observations	1,013	1,013	1,013	1,013	1,013
Adj./Ps. R^2	0.04	0.02	0.01	0.01	0.02

recalled more orange than blue numbers in the *blue* treatment (30.4% to 44.2%, significant at 1% level). Across Columns (1) and (2), the treatment dummy coefficients are close. In fact, answers in Q1, Q3 and Q4 are consistent for roughly 90% of participants. Columns (3) and (4) show how the median quantity of recalled orange and blue numbers depends on the treatment. Responses in the *gray* treatment are quite accurate, as indicated by the constant term. In the *blue* treatment, participants retrieve fewer blue numbers, consistent with interference from blue words.

Finally, based on answers to Q3 and Q4, we can compute the ratio of orange numbers to total numbers recalled. Column (5) shows that, as predicted by the model, participants recalled on average a significantly higher share of orange numbers in the *blue* treatment

(50% versus 44.5%). Given that the recalled share of orange numbers is high at baseline, this average increase can have a large effect on the number of participants who say orange is the likely color.

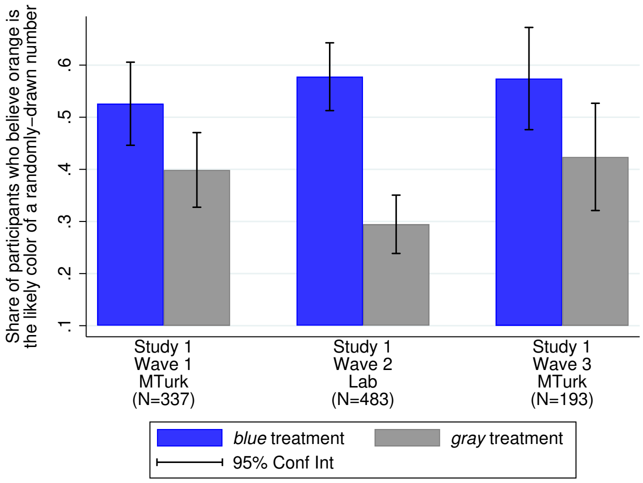


Figure B.6: Share of participants who believe that the likely color of a randomly-drawn number is orange for the *blue* and *gray* treatments for each wave of Study 1b.

Result per wave Figure B.5 and Table B.4 presented the main findings of Study 1b on the outcomes Q1, Q3, and Q4 which provided support for the predictions of our model.

Figure B.6 shows that the share of participants who recalled more orange than blue numbers is larger in the *blue* treatment than in the *gray* treatment for each wave. These differences of 12.71pp, 28.32pp, and 15.03pp for waves 1, 2, and 3, respectively, are significant in OLS regressions (p-values of 0.020, <0.001, and 0.037, respectively). Note that the treatment effect for wave 2—which was conducted in the laboratory—is significantly larger than the pooled treatment effect of waves 1 and 3—which were conducted with MTurk—in a OLS difference-in-differences regression (p-value of 0.018 for the difference-in-differences estimate). Thus, in the laboratory we find a greater treatment effect

Tables B.5, B.6, and B.7 show that the results of Table B.4 also hold when focusing only on wave 1, wave 2, or wave 3, respectively. Only the treatment effects on the median orange numbers recalled and median blue numbers recalled do not replicate for wave 3, which has the smallest sample of the three waves.

Different tests We show in Table B.8 that our results presented in Table B.4 are robust to using different statistical tests (Logit regressions instead of OLS regressions for outcome measures of Columns 1 and 2 as well as OLS regressions instead of 0.5 quantile

Table B.5: Regression estimates of treatment effects in Study 1b for wave 1

	OLS: Y=1 if "orange is likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.1271**	.0758	0	− 2**	.0556***
	(.0542)	(.0527)	(.3916)	(.9791)	(.0100)
Constant	.3989***	.3333***	10***	14***	.4444***
	(.0366)	(.0356)	(.2647)	(.6618)	(.0144)
Observations	337	337	337	337	337
Adj./Ps. R^2	0.02	0.01	0.00	0.01	0.02

Table B.6: Regression estimates of treatment effects in Study 1b for wave 2

	OLS: Y=1 if "orange is likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.2832***	.1754	0	− 2***	.0556***
	(.0433)	(.0434)	(.4541)	(.7241)	(.0175)
Constant	.2946***	.2868***	10***	14***	.4444***
	(.0366)	(.0296)	(.3099)	(.4942)	(.0119)
Observations	483	483	483	483	483
Adj./Ps. R^2	0.08	0.01	0.00	0.02	0.02

regressions for Columns 3, 4, and 5). Participants are significantly more likely to believe that a randomly-drawn image is likely to be orange (Column 1). Participants are significantly more likely to recall more orange than blue numbers in the *blue* treatment than in the *gray* treatment (Column 2). Additionally, participants state a significantly greater average share of orange numbers recalled to total amount of images recalled in the *blue* treatment than in the *gray* treatment (Column 5). Columns 3 and 4 show that subjects recall on average more orange numbers and less blue numbers in the the *blue* treatment than in the *gray* treatment, however these differences are not significant.

Table B.7: Regression estimates of treatment effects in Study 1b for wave 3

	OLS: Y=1 if “orange is likely”	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.1503**	.1521**	2	3**	.0556**
	(.0716)	(.0692)	(1.612)	(1.408)	(.0100)
Constant	.4239***	.2935***	10***	12***	.4444***
	(.0518)	(.0501)	(1.166)	(1.019)	(.0190)
Observations	193	193	193	193	193
Adj./Ps. R^2	0.02	0.02	0.00	0.01	0.03

Table B.8: Robustness of regression estimates of treatment effects in Study 1

	Logit: Y=1 if “orange is likely”	Logit: Y= 1 if more orange numbers recalled	OLS Y= Orange numbers recalled	OLS Y= Blue numbers recalled	OLS Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.8450***	.5948***	.4667	− .1399	.0341***
	(.0307)	(.0302)	(.4130)	(.5190)	(.0100)
MTurk dummy	yes	yes	yes	yes	yes
Wave dummies	yes	yes	yes	yes	yes
Constant	−.7022***	−.8270***	11.56***	14.06***	.4517***
	(.1137)	(.1162)	(.3549)	(.4461)	(.0144)
Observations	1,013	1,013	1,013	1,013	1,013
Adj./Ps. R^2	0.03	0.02	0.01	0.00	0.01

B.1.2 Differences between Study 1 and Study 1b

Table B.9 shows OLS difference-in-difference regressions that test whether the treatment effects of Study 1 (with distraction) and Study 1b (without distraction) differ significantly from each other. The highlighted row shows the difference-in-differences estimates, which are zero or close to zero for all dependent variables in size and do not differ from zero significantly for any of the dependent variables.

Table B.9: Comparing regression estimates of treatment effects between Studies 1 & 1b

	OLS: Y=1 if "orange is more likely"	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg Y= Orange numbers recalled	0.5-Q-Reg Y= Blue numbers recalled	0.5-Q-Reg: Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *blue*	.2060***	.1379***	0	− 2***	.0556***
	(.0307)	(.0302)	(.4187)	(.6427)	(.0124)
1 if distraction	− .0549	− .0821**	0	0	−.0159
	(.0409)	(.0398)	(.6253)	(.8114)	(.0164)
1 if *blue* & distraction	**.0051**	**.0288**	**0**	**0**	**0**
	(.0460)	**(.0448)**	**(.7041)**	**(.9134)**	**(.0185)**
MTurk dummy	yes	yes	yes	yes	yes
Wave dummies	yes	yes	yes	yes	yes
Constant	.3305***	.3043***	10***	14***	.4444***
	(.0302)	(.0259)	(.3598)	(.5524)	(.0107)
Observations	1,803	1,803	1,803	1,803	1,803
Adj./Ps. R^2	0.05	0.02	0.01	0.01	0.02

B.2 Study 2

Orange words

Study 2 implements the same design as Study 1 *blue words* treatment, with the one difference that some of the words in the decoy group were shown in orange and the remaining were shown in blue as before.

Procedural details and data collection

We conducted three waves of Study 2's treatments. The first two waves were conducted in March and May of 2018 with MTurk, $N = 307$ and $N = 1,431$, respectively. In both of these waves, the experiment only consisted of the blue treatment variation with orange words. The experiment lasted for 7 minutes for both waves. Participants received a $1.00 show-up fee as well as $0.20 for each correct answer. In the first wave we conducted the treatments with $k = 0, 1, 3, 6$. In the second wave we replicated the first wave and included in addition treatments with $k = 10, 22$.

The third wave was conducted in May of 2018 in the laboratory of Bocconi University with $k = 1, 6$. These treatments were accompanied by an unrelated intertemporal choice—like in the laboratory experiments of Studies 1 and 1b. The entire lab experiment took 10 minutes. Subjects received a show-up fee of €4.00. In case they were randomly selected to receive additional payments based on the *blue* treatments with

$k = 1, 6$, subjects received €0.50 for each correct answer to the 50 images. We are using the *blue* treatment of the lab experiment of Study 1 and Study 1b as a comparison standard.

Further results

Table B.10 shows our findings when looking at MTurk participants only. Like in Table 4 from the main text that shows our results when looking at lab participants only, we find evidence for Prediction 3. The share of participants who say "orange is more likely" drops by 1pp per orange word added. Turning to data on recall, as orange words are added, participants recall more blue numbers and are less likely to recall seeing more orange numbers.

Table B.10: Regression estimates of treatment effects in Study 2 (MTurk only)

	OLS: Y=1 if "orange is likely"	0.5-Q-Reg: Y= Orange numbers recalled	0.5-Q-Reg: Y= Blue numbers recalled	OLS: Y= 1 if more orange numbers recalled	0.5-Q-Reg: Y= Share of orange to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
k (number of orange words)	−.0093***	0	.0625*	−.0030**	−.0005
	(.0015)	(.0488)	(.0334)	(.0015)	(.0006)
Constant	.5265***	12***	13.625***	.4116***	.5005***
	(.0015)	(.5031)	(.3421)	(.0156)	(.0058)
Observations	1,839	1,839	1,839	1,839	1,839
Adj./Ps. R^2	0.02	0.00	0.00	0.02	0.00

Notes: This table presents estimates of the treatment effect on several outcome measures relating to our model predictions in Study 2. We report on our results from MTurk here. Columns (1) and (5) present OLS regression of response dummies (1 if "orange is likely" in Column (1) and 1 if "recalled more orange than blue numbers" in Column (3)) on k, the number of orange words. Columns (2), (3), (4) and (6) present 0.5-quantile regressions of subjects' stated probability that a random number is orange, number of recalled orange numbers, number of recalled blue numbers and share of recalled oranges numbers, respectively, on k. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

B.3 Study 3

Questions

The following questions were asked to participants, in the order presented below. The labelling used here mirrors that in the main text.

Color Cue Treatment

- Screen 1
  - Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely color of the chosen image? Blue or Orange.
- Screen 2
  - Q2: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the probability that this number is orange?
- Screen 3
  - Q3a: How many blue numbers in small font size were shown to you?
  - Q3b: How many blue numbers in large font size were shown to you?
  - Q4a: How many orange numbers in small font size were shown to you?
  - Q4b: How many orange numbers in large font size were shown to you?

Size Cue Treatment (per screen in order of display)

- Screen 1
  - Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely font size of the chosen image? Small or Large.
- Screen 2
  - Q4: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the probability that this number is large?
- Screen 3

- Q2a: How many blue numbers in small font size were shown to you?
- Q2b: How many blue numbers in large font size were shown to you?
- Q3a: How many orange numbers in small font size were shown to you?
- Q3b: How many orange numbers in large font size were shown to you?

Procedural details and data collection

We conducted two waves of Study 3. Wave 1 was conducted in May of 2018 in the laboratory of the Bocconi University ($N = 326$). Wave 2 replicated Wave 1 in October of 2018 in the laboratory of the University of Cologne ($N = 321$). The memory treatments were accompanied by an unrelated intertemporal choice—like in the laboratory experiment of Studies 1 and 1b. The entire lab experiment took 10 minutes. Participants received a show-up fee of €4.00. In case they were randomly selected to receive additional payments based on the treatments of Study 3, participants received €0.50 for each correct answer to the questions Q1-Q4.

Further results

We find evidence for our prediction for both Q1 and Q4 in both waves of Study 3, see Figures B.7 and B.8.

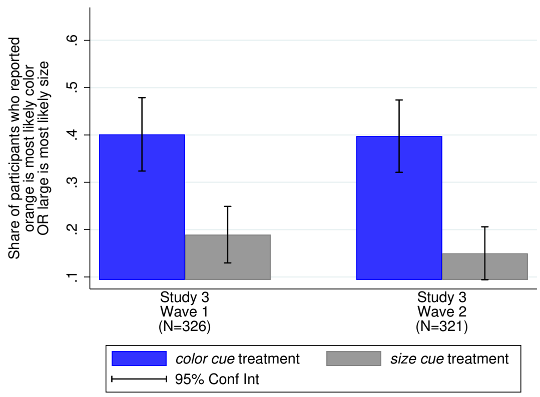


Figure B.7: Share of participants who believe that the color OR size of a randomly-drawn number is most likely orange OR large for the treatments of Study 3.

B.3.1 Within-subject comparison of outcome measures across Studies 1, 2, 3

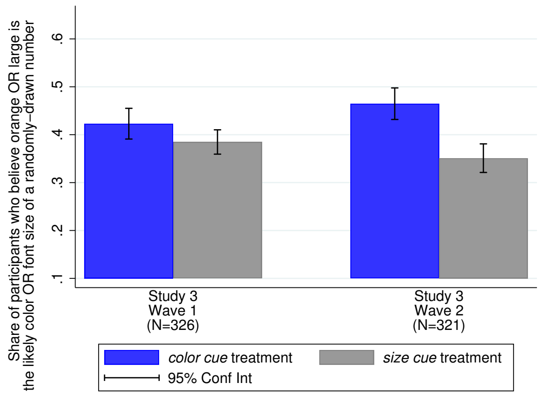


Figure B.8: Participants' belief that a random number is orange OR large for the treatments of Study 3.

Table B.11: Relationship between the share of recalled orange over total numbers and the direct probability measure

	OLS: Y= Probability that a randomly-drawn number is orange		
	Study 1	Study 2	Study 3
	(1)	(2)	(3)
Share of recalled orange over total numbers	.8607*** (.0278)	.7284*** (.0628)	.3751*** (.0432)
Wave dummy	yes		yes
Constant	.0611*** (.0139)	.1272*** (.0307)	.2410*** (.0204)
Observations	790	254	644
Adj./Ps. R^2	0.55	0.35	0.10

Notes: This table presents estimates of the relationship between the share of recalled orange over total numbers and the direct probability measure. Columns (1), (2) and (3) present OLS regression of participants' imputed stated probability that a random number is orange on their stated share of recalled orange over total numbers, respectively, for each of study 1, study 2 and study 3. Standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

C Further Experiments

In the following, we discuss 4 further studies that we conducted. The first study, which we call study 1.A in the main text, builds on study 1 but varies the number of decoy images presented. This study is used in the assessment of the performance of the calibrated model in Section 5. The second study instead varies the composition of the target group. The third and fourth studies vary other design features.

C.1 Study 1.A

Design

In Study 1.A, we vary the number of decoy images, as well as their content (blue words vs gray shapes), in a 4 x 2 design presented in Table C.1.

Table C.1: Study 1.A's *Blue j* Treatment and *Gray j* Treatment

***Blue j* Treatments**
(with $j \in \{5, 50, 75, 125\}$)

	Orange	Blue	Gray
Target data	10 Orange Numbers	15 Blue Numbers	
Decoy data		j Blue Words	

***Gray j* Treatments**
(with $j \in \{5, 50, 75, 125\}$)

	Orange	Blue	Gray
Target data	10 Orange Numbers	15 Blue Numbers	
Decoy data			j Gray Shapes

Questions The following questions were asked to participants, in the order presented below.

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely color of the chosen image? Blue or Orange.
- Q2: How many orange numbers were shown to you?
- Q3: How many blue numbers were shown to you?
- *Blue* treatments only:

AddQ1: How many blue words were shown to you?

AddQ2: How many orange words were shown to you?

- *Gray* treatments only:

AddQ1: How many gray shapes were shown to you?

Procedural details and data collection

We conducted two waves of Study 1.A. The first wave was conducted in March of 2018 with MTurk and a sample of 800 participants featuring all treatment cells of the 4 × 2 design. We then replicated the *blue* treatments with $j = 5$, 50, and 75 decoys in May of 2018 with MTurk and a sample of 592 participants. In both waves, the experiment consisted only of our memory treatments. The experiment lasted for 7 minutes. Participants received a $1.00 show-up fee as well as $0.20 for each correct answer to the questions on the 50 images.

Results

Figure C.1 shows that we find no treatment effect when the blue and orange numbers are displayed along 5 decoys (in the *blue* and the *gray* treatments). For 50, 75, and 125 decoys, however, we do find significant treatment effects that resemble our findings of Study 1.

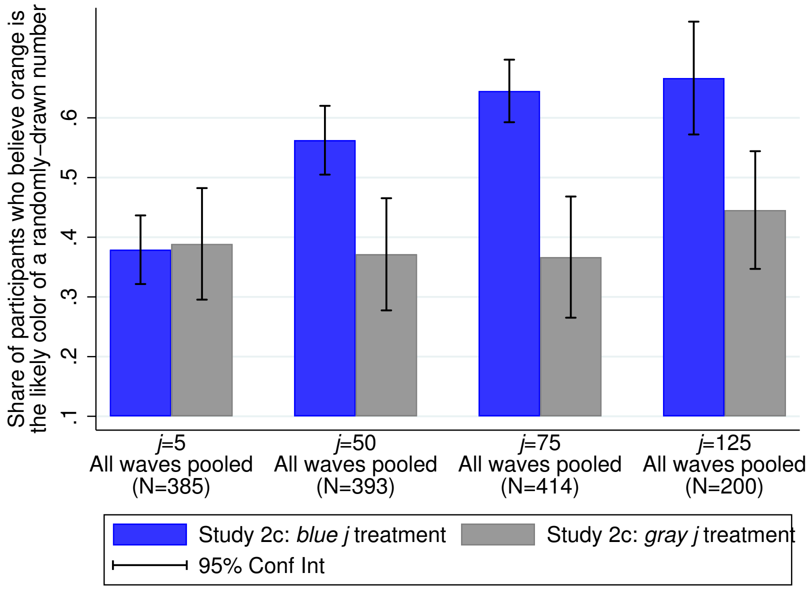


Figure C.1: Share of participants who state orange is the likely color, by treatment.

C.2 Study 2.A

Design

In Study 2.A, we vary likelihood ratios by changing the composition of images within our target data, as well as their content (blue words vs gray shapes), in a 3 x 2 design presented in Table C.2.

Table C.2: Study 2.A's *Blue i* Treatment and *Gray i* Treatment

Blue i Treatments
(with **i** $\in \{13, 20, 25\}$)

	Orange	Blue	Gray
Target data	25—**i** Orange Numbers	**i** Blue Numbers	
Decoy data		25 Blue Words	

Gray i Treatments
(with **i** $\in \{13, 20, 25\}$)

	Orange	Blue	Gray
Target data	25—**i** Orange Numbers	**i** Blue Numbers	
Decoy data			25 Gray Shapes

Questions The following questions were asked to participants, in the order presented below.

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely color of the chosen image? Blue or Orange.
- Q2: How many orange numbers were shown to you?
- Q3: How many blue numbers were shown to you?
- *Blue* treatments only:

AddQ1: How many blue words were shown to you?

AddQ2: How many orange words were shown to you?

- *Gray* treatments only:

AddQ1: How many gray shapes were shown to you?

Predictions

Our prediction is that as we increase the share of blue numbers, the likelihood that participants recall a randomly-chosen number as orange should decrease in the *blue* treatment. Thus, we expect a smaller treatment effect (difference between *blue* and *gray*) as the share of blue numbers increases. Denote the number of blue numbers as i, so that the number of orange numbers is 25-i and assume $c > 0$. Then, representativeness-based recall yields the following prediction:

As the share of blue to orange numbers increases, the share of participants who believe that the likely color of a random number is orange should decrease in the *blue* treatment, because the assessed probability that a random number is blue increases, formally $\tilde{P}(b|n)_{blue\ i} \geq \tilde{P}(b|n)_{blue\ i'}$ for $i > i'$.

Procedural details and data collection

We conducted two waves of Study 2.A. The first wave was conducted in March of 2018 via MTurk with 601 participants. The experiment consisted only of our memory treatments. The experiment lasted for 7 minutes. Participants received a $1.00 show-up fee as well as $0.20 for each correct answer. We then replicated these results in the laboratory of the University of Cologne with 516 participants. The memory treatments were accompanied by an unrelated intertemporal choice—like in the laboratory experiment of Study 1. The entire lab experiment took 10 minutes. Participants received a show-up fee of €4.00. In case they were randomly selected to receive additional payments based on our treatments on memory and representativeness, participants received €0.50 for each correct answer to the questions on the 50 images.

Results

Our findings provide evidence for our prediction. When the share of blue to orange numbers is increased, participants' *blue* to *gray* treatment difference in their assessed probability that a random number is orange decreases. The treatment effect on the share of participants believing that "orange is more likely" is 15.3pp for 13 blue numbers and 12 orange numbers. The treatment effect reduces to 10pp for 20 blue and 5 orange numbers and to 6pp for 23 blue and 2 orange numbers. While all three treatment effects are (at least weakly) significantly different from zero, the former one is larger than the latter two. OLS regressions show that this difference in treatment effects is at most weakly significantly different from zero when comparing the 13 blue numbers and 12 orange numbers case with the 23 blue numbers and 2 orange numbers case as well as when comparing the 13 blue numbers and 12 orange numbers case with the pooled cases of 23 blue numbers and 2 orange numbers as well as 20 blue numbers and 5 orange numbers. Column (1) of Table C.3 shows the results of an OLS regression of the latter result. The weakly significant interaction term (Row (3)) implies the discussed difference in treatment effects. Column (2) of Table C.3 shows that the difference in treatment effects increases in size and significance when the main treatments of Study 1 are included:

The treatment effect for treatments with 20 or 23 blue numbers is significantly smaller than for treatments with 13 or 15 blue numbers.

Table C.3: Regression estimates of treatment effects in Study 2.A

	Including only Study 2.A OLS: Y=1 if "Orange is more likely"	Including Study 1 & 2.A OLS: Y=1 if "Orange is more likely"
	(1)	(2)
1 if *blue*	.1669*** (.0411)	.1957*** (.0237)
1 if i (Blue Numbers) ≥20	−.3044*** (.0354)	−.2897*** (.0397)
1 if *blue* & $i \geq 20$	−.0904* (.0499)	−.1192*** (.0656)
MTurk dummy	yes	yes
wave dummies	–	yes
Constant	.3302*** (.0325)	.3153*** (.0300)
Observations	1,117	2,130
Adj. R^2	0.18	0.15

Notes: For Study 2.A, wave dummy and MTurk dummy are collinear

C.3 Fontsize Study

The design of our fontsize Study is similar to that of Study 1, except that images vary in font size as opposed to color.

Design

The two between-subjects treatments of our fontsize Study follow the same structure as our baseline treatments of Study 1:

First, participants are told they will see questions that are incentivized for accuracy about a sequence of 50 abstract images displayed to them during the experiment.

Second, participants see the 50 images, each of which appears on separate screens for short moments of time and in random order. The 50 images vary along two features. As in previous Studies, images can be a number or a word. We continue to refer to numbers as targets and words as decoys. The second feature is the font size of the object, which

Table C.4: Fontsize Study's *small* treatment and *large* treatment

Small **(Font Size) Treatment**

	Large	Small
Target data	10 Large Numbers	15 Small Numbers
Decoy data		25 Small Words

Large **(Font Size) Treatment**

	Large	Small
Target data	10 Large Numbers	15 Small Numbers
Decoy data	25 Large Numbers	

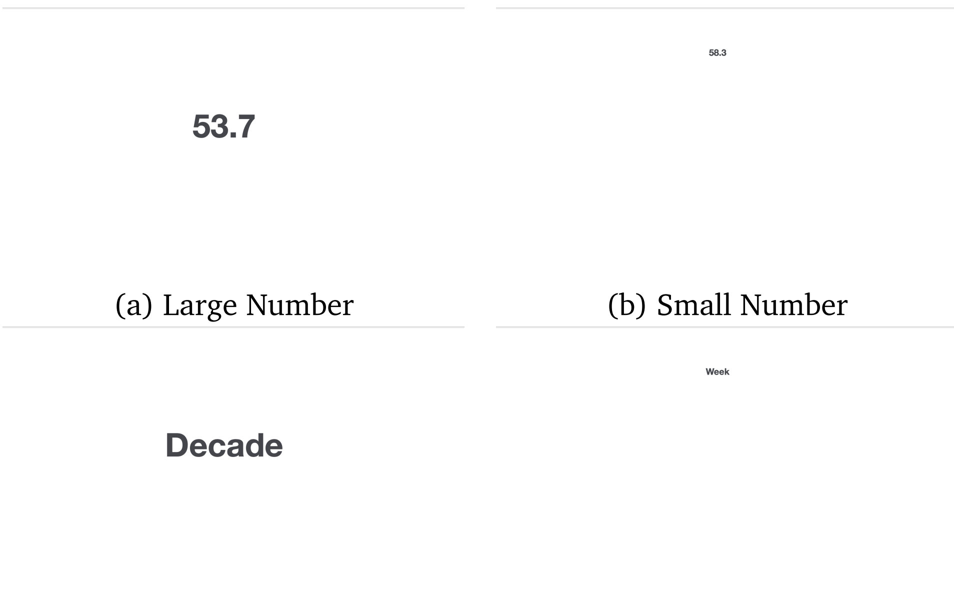
can be large or small. Table C.4 shows which types of images participants were shown for each of the two treatments. Example screenshots for each kind of image are displayed in Figure C.2.

Third, participants face questions on the targets which require them to recall the observed sequence of images as well as on the decoys images. The questions – presented here in the same order as shown to participants in the experiment—were:

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a number. What is the likely size of the chosen image? Large or Small.
- Q2: How many large numbers were shown to you?
- Q3: How many small numbers were shown to you?
- Q4: How many small words were shown to you?
- Q5: How many large words were shown to you?

Predictions

Our prediction is that as we change the decoy images from being large words to small words, the degree of participants who believe that a randomly-chosen number was shown in large font size should be greater. Because large numbers are representative only in the *small* treatment, interference-based recall inhibits recall of small numbers more in the *small* treatment than in the the *large* treatment. Thus, we expect more participants to state that the likely font size is large in the *small* treatment than in the *large* treatment.



(a) Large Number (b) Small Number

(c) Large Word (d) Small Word

Figure C.2: Examples of images shown to participants

Procedural details and data collection

We conducted the fontsize Study in December of 2017 with MTurk and a sample of 374 participants. Our *large* and *small* treatments were accompanied by unrelated intertemporal choices. The entire experiment lasted for around 13 minutes. Participants received a $1.00 show-up fee. A computer-based coin toss determined randomly whether subjects would receive additional payments based on the *blue* and *gray* treatments or on the unrelated intertemporal choice. In case subjects received additional payments based on the former, one of every 100 participants was randomly chosen to receive $20.00 for each correct answer to Q1-Q5, while all remaining participants received $1 for each correct answer.

Results

We find that 22.4% of participants believe that the likely font size of a randomly-drawn number is large in the *large* treatment. However, in the *small* treatment, 32% of participants believe that the likely font size of a randomly-drawn number is large. This difference is consistent with our model's prediction and is significantly greater than zero in an OLS regression, see Table C.5 Column 1.

Mirroring our results from Study 1, we also find that a greater share of participants recalls more large than small numbers in the *small* treatment than in the *large* treatment, Column 2 of Table C.5. However, we also find that the median amount of recalled large numbers as well as the median amount of recalled small numbers do not differ across treatments.

Table C.5: Regression estimates of treatment effects in the fontsize Sudy

	OLS: Y=1 if "orange is likely"	OLS: Y= 1 if more large numbers recalled	0.5-Q-Reg Y= Large numbers recalled	0.5-Q-Reg Y= Small numbers recalled	0.5-Q-Reg Y= Share of large to total numbers recalled
	(1)	(2)	(3)	(4)	(5)
1 if *small*	.1006**	.0883**	0	0	.0274*
	(.0460)	(.0402)	(.8609)	(.9589)	(.0157)
Constant	.2240***	.1421***	8***	15***	.3571***
	(.0329)	(.0287)	(.6152)	(.6852)	(.0112)
Observations	374	374	374	374	373
Adj./Ps. R^2	0.01	0.01	0.00	0.00	0.00

C.4 Study ProbNumber

Design

The two between-subjects treatments of our Study ProbNumber follow the same structure as our baseline treatments of Study 1:

First, participants are told they will be asked questions that are incentivized for accuracy about a sequence of 50 abstract images displayed to them during the experiment.

Second, participants see the 50 images that appear on separate screens for short moments of time and in random order. The 50 images vary along two features. The first feature is the object category: each object can be a number, a word or shape. The second feature is color, which can be blue or orange. Table C.6 shows which types of images participants were exposed for each of the two treatments. Example screenshots for each kind of image are displayed in Figure C.3. In the ProbNumber study, blue objects are the target images and the orange objects are the decoys.

Third, participants face one question on the targets which require them to recall the observed sequence of images:

- Q1: The computer randomly chose 1 image from all images that were just shown to you. The chosen image showed a blue object. What is the probability that the chosen image is a number?

Table C.6: Study ProbNumber's *numbers* treatment and *shapes* treatment

***Numbers* Treatment**

	Decoy data	Target data
Numbers	10 Orange Numbers	15 Blue Numbers
Non-numbers		25 Blue Words

Shapes Treatment

	Decoy data	Large data
Numbers		15 Blue Numbers
Non-numbers	10 Orange Shapes	25 Blue Words

Prediction

Our prediction is that as we change the decoy images from being orange shapes to orange numbers, the likelihood that participants recall a randomly-chosen blue object as being a number should be lower. Because orange numbers are representative only in the *numbers* treatment, interference-based recall inhibits recall of blue numbers more in the *numbers* treatment than in the the *shapes* treatment. Thus, we expect participants to state lower probabilities that a randomly-chosen blue object is a number in the *numbers* treatment than in the *shapes* treatment.

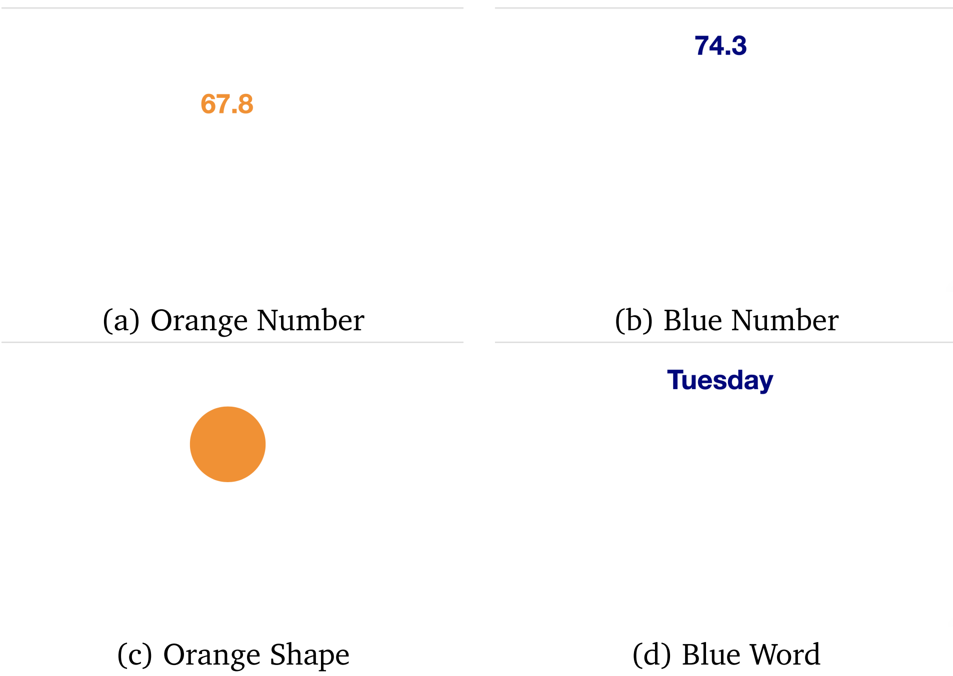
Procedural details and data collection

We conducted the ProbNumber Study in March of 2018 with MTurk and a sample of 304 participants. The experiment consisted only of the *numbers* and *shapes* treatments. The entire experiment lasted for around 7 minutes. Participants received a $1.00 show-up fee and $1.00 for a correct answer to Q1.

Results

Consistent with our prediction, we find that participants believe that a randomly-drawn blue object is a number to a greater extend in the *shapes* treatment—where orange shapes are unlikely to interfere with the recall of blue numbers—than in the *numbers* treatment—where orange numbers are predicted to interfere with the recall of blue numbers.

The median probability that a blue object is a number is 50% in the *shapes* treatment and 40% in the *numbers* treatment. This difference is significant at the 1% level in a



(a) Orange Number (b) Blue Number

(c) Orange Shape (d) Blue Word

Figure C.3: Examples of images shown to participants in Study ProbNumbers

0.5 quantile regression. We find a similar treatment effect when looking at the mean instead of the median. The mean probability that a blue object is a number is 48% in the *shapes* treatment and 42% in the *numbers* treatment. This difference is significant at the 5% level in an OLS regression.

The results of ProbNumber provide further evidence for how interference-based recall drives distortions of probabilistic judgements.

D Calibration

D.1 Model parameters

We calibrate the model parameters α, γ, c by exploring a grid of parameter space to minimize the least squared error

$$LSE(\alpha,\gamma,c) = \sum_i \left(\frac{\tilde{P}(o|n)_i}{\tilde{P}(b|n)_i} - \frac{P(o|n)_i}{P(b|n)_i} \right)^2$$

where $\frac{\tilde{P}(o|n)_i}{\tilde{P}(b|n)_i}$ are the predicted moments in Equation (6), which depend on the parameters, and $\frac{P(o|n)_i}{P(b|n)_i}$ are the target moments in Table 6. We use a grid method because Equation 6 is very non-linear for c close to zero, which makes standard numerical solvers unstable. The grid resolution is 0.01 for α and γ, and 0.005 for c, and covers the range$[0,2]_\alpha \times [-0.2, 0.2]_\gamma \times [0, 0.2]_c$. The reported calibrated parameters α^*, γ^*, c^* are the average of the ten grid elements with lowest LSE. As illustrated in Figure 3, given the target moments these parameters are tightly estimated.

To assess statistical significance of the key model parameter γ, we generate a distribution of parameters matched to a large bootstrapped set of target moments, as follows. For experiments $k = 0$ and $k = 1$ in Table 6, we draw (with replacement) N observations from the actual distribution of assessments of $P(o|n)$, where N is the sample size in the experiment. Repeating this procedure a large number of times yields a (marginal) distribution for each target moment. We then estimate the distribution of γ to match the distribution of moments, keeping $\alpha = \alpha^*$ and $c = c^*$. Keeping $c = c^*$ allows us to use fast numerical solvers to obtain distributions of α and γ in reasonable timeframes. As discussed in footnote 21, we do not re-estimate c because it has little influence in the target moments other than moderating the effect of representativeness of orange numbers when there are no orange words, and we know that the best-fitting c is positive. We obtain the following 95% confidence interval: $CI_\gamma = [0.05, 0.086]$.

Similarly, we estimate the distribution of α to match the distribution of moments, keeping $\gamma = \gamma^*$ and $c = c^*$, and obtain the 95% confidence interval: $CI_\alpha = [0.47, 0.84]$.

D.2 Out of sample predictions

To assess performance of the model against benchmark of bayesian beliefs with noise, we take the distribution of de-meaned forecast errors for each of the three target experiments in Table 6, i.e. the distribution of $\widehat{P}_{i,exp}(o|n) - \frac{1}{N_{exp}} \sum_i \widehat{P}_{exp}(o|n)$ where i indexes participants, exp indexes experiments, and N_{exp} is the corresponding sample size. Call the resulting distribution D_{exp}.

We can now assess whether a benchmark of bayesian beliefs with noise given by this distribution can account for our results. For each target experiment, we draw N_{exp} observations from D_{exp} with replacement. For each draw we obtain a pseudo-experimental distribution of answers. We then compute the distribution of sample mean errors for a large number of draws (10,000). This yields a confidence interval around the experimental results. A similar procedue, where we pool the de-meaned forecast errors from all experiments, yields a confidence interval around the true (null) value, $P(o|n) = 0.4$.

We present Figures 4, 5, 6 and 7, adding the corresponding confidence bands. We also include confidence intervals on the model predictions that follow from the confidence interval on parameter γ estimated above.

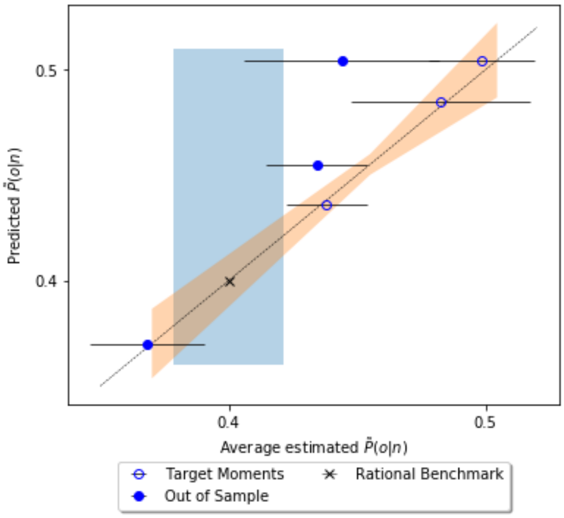


Figure D.4: A plot of Figure 4 with confidence intervals. The shaded blue band represents the confidence interval for the rational benchmark model with noise. The shaded orange band represents the confidence interval for our model's predictions, accounting for uncertainty about parameter γ.

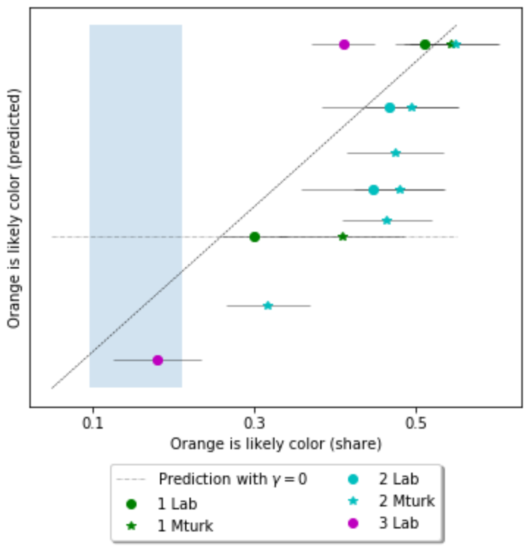


Figure D.5: A plot of Figure 5 with confidence intervals. The shaded band represents the confidence interval for the rational benchmark model with noise.

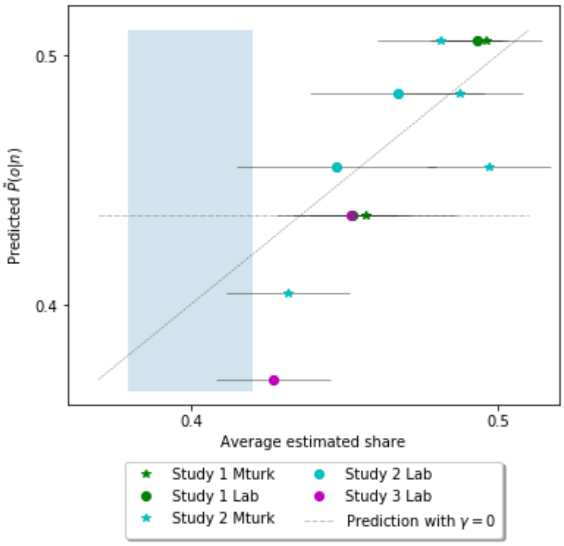


Figure D.6: A plot of Figure 6 with confidence intervals. The shaded band represents the confidence interval for the rational benchmark model with noise.

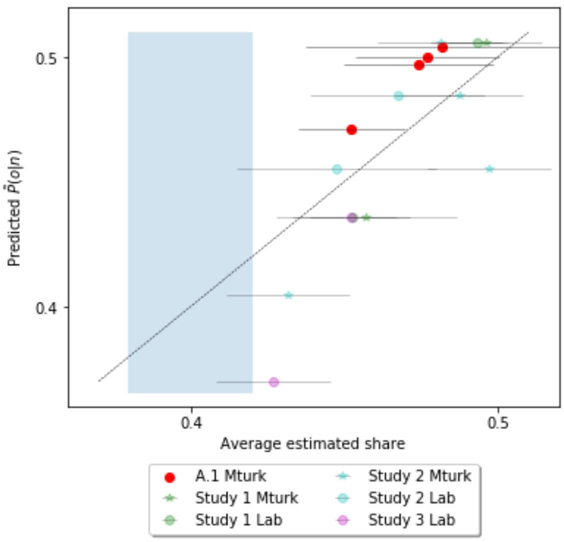


Figure D.7: A plot of Figure 7 with confidence intervals. The shaded band represents the confidence interval for the rational benchmark model with noise.