AMERICAN PSYCHOLOGICAL ASSOCIATION

Decision

ISSN: 2325-9965

2025, Vol. 12, No. 1, 31–42
https://doi.org/10.1037/dec0000254

BRIEF REPORT

Visual Simulation and Subjective Probability Estimation: When Seeing Is Believing

Samuel N. Kirshner
UNSW Business School, University of New South Wales

Despite the volume of research examining overprecision, the underlying drivers of individual differences in probability estimations remain elusive. I propose that visual processing through mental simulation of small samples is a cognitive mechanism influencing the relative degrees of over and underprecision. I conducted three preregistered experiments contrasting probability estimates between a control group and a treatment group, where participants were prompted to engage in greater visual processing by mentally simulating outcomes. In Study 1, participants estimate a binomial distribution of a ball drop machine. I find that engaging in visual simulation led to higher estimates of values near the distributions' center, while the control group provided higher estimates for the distributions' tails. Although the control group is more underprecise, visual simulation could arguably increase estimation accuracy and not overconfidence. To separate these effects, I modify the ball drop mechanism in Study 2 to produce a flatter distribution. The results show that the control group is well-calibrated, but the visual simulation group is overprecise. Study 3 investigates a boundary condition where participants mentally simulated multiple outcomes. The results demonstrate that an increase in the variance of imagined outcomes lowers subjective estimates near the center of the distribution, diminishing the treatment effect.

Keywords: overprecision, visual processing, mental simulation, subjective probability estimates, overconfidence

Supplemental materials: https://doi.org/10.1037/dec0000254.supp

This article was published Online First November 7, 2024.
Samuel N. Kirshner https://orcid.org/0000-0002-2604-2001

The experiments in this study were approved by University of New South Wales's institutional review board, and all participants provided informed consent before participating in the study. The experimental materials, preregistration, data, and R code for the analysis are available on the Open Science Framework at https://osf.io/4j5vy/?view_only=3295e2c8518e43259ad8caef20dfc9e2. If there are any questions or comments, kindly contact Samuel N. Kirshner.

The author declares that there are no conflicts of interest associated with this research. The author has reported all implemented experimental conditions and disclosed all measured variables. The author also declares that he have reported all studies run on the research question of the article. The author thanks the UNSW Business School for funding this research.

Open Access funding provided by University of New South Wales:

Samuel N. Kirshner played a lead role in conceptualization, data curation, formal analysis, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Samuel N. Kirshner, UNSW Business School, University of New South Wales, 2101 Quadrangle Building, UNSW Kensington Campus, Sydney, NSW 2052, Australia. Email: s.kirshner@unsw.edu.au

Many individual, organizational, and societal decision problems require judgments on likelihoods and probability distributions (Fellner & Krügel, 2012). However, substantial empirical evidence demonstrates that individuals poorly interpret and estimate probability distributions (e.g., Hogarth, 1975). Overconfidence is a prominent driver of poorly calibrated probability estimates, with many researchers systematically finding that judgments are overprecise, that is, display excessive confidence in the accuracy of beliefs (Moore & Dev, 2020), which is one of three types of overconfidence (Moore, 2023). Yet, cognitive mechanisms and individual differences driving overprecision remain elusive despite the volume of research on overconfidence (Moore & Dev, 2020; Moore et al., 2015).

The literature primarily examines overprecision characterized by epistemic uncertainty, knowable events that do not have calculable distributions. These studies ask participants to provide a 90% confidence interval for knowledge questions (e.g., how long is the Nile River) and examine if 90% of the answers fall within the provided ranges. The 90% confidence intervals consistently contain the actual value less than 50% of the time, an effect that Moore et al. (2015) noted "is one of the most dramatic and impressive in the decision-making literature and has been replicated in many paradigms and populations" (p. 110).

However, without a normative distribution for comparison, measuring overprecision using epistemic uncertainty obfuscates whether prediction intervals are too narrow compared to the objective distributions or whether estimates are centered on incorrect means. Accordingly, Moore et al. (2015) investigate if individuals are overprecise in their estimations due to overly narrow confidence intervals by considering problems with aleatory uncertainty, that is, events with knowable probability distributions. Remarkably, Moore et al. (2015) found that their "respondents are catastrophically underprecise" (p. 112),[1] providing direct evidence against existing theories explaining subjective probability estimates, such as anchor and insufficient adjustment (Tversky & Kahneman, 1974) and naive statistics (Juslin et al., 2007), which predict overprecision. Thus, despite the extensive literature on overconfidence, little is known of the cognitive drivers of how individuals construct over and underprecise distributions (Moore et al., 2015).

This article proposes that an individual's propensity to mentally simulate outcomes provides a systematic driver for relative differences in over and underprecision. Research examining the availability heuristic shows that individuals who imagine events more readily form expectations of those events occurring. For example, Carroll (1978) found that subjects imagining Jimmy Carter winning the 1976 presidential election were likelier to predict that Jimmy Carter would win than subjects imagining Gerald Ford as the victor. Anderson (1983) found that people who imagined themselves undertaking an event and drew a cartoon of the event (e.g., applying for a part-time job) had greater estimates of the event's behavioral intentions (e.g., the likelihood of applying for a part-time job) compared to imagining others. Similarly, many other studies provide empirical evidence that imagining outcomes influences beliefs that outcomes would occur (Gregory et al., 1982; Sherman et al., 1981, 1985). More recently, Bilgin (2012) extended imaging outcomes to prospect theory, supporting that mentally simulating prospects can explain the asymmetry between losses and gains. Attari et al. (2022) showed that observing outcomes leads to overestimating the likelihood of an event occurring. Yet, the extant literature exclusively studies the impact of mental simulation on events and scenarios with binary outcomes (e.g., Jimmy Carter will win/lose the election; apply/do not apply for a part-time job; accept/reject a gamble).

To explore the potential impact of mental simulation on probability estimates, I leveraged the theory of naïve statistics of mental samples (Juslin et al., 2007; Tong & Feiler, 2017) and the "2-error" hypothesis (Marchiori et al., 2015). If an

[1] Moore et al. (2015) considered both traditional hit rates and Subjective Probability Interval Estimate (SPIES), which force participants to choose probabilities for given intervals that cover the full range of potential outcomes (Haran et al., 2010). For aleatory uncertainty, participants' distributions were too wide, providing underprecision, but they were not centered around the true mean, leading to overprecision on hit rates. There are also methodological concerns with measuring overprecision using interval production tasks. For example, Langnickel and Zeisberger (2016) experimentally showed that decision makers' interval confidence for general knowledge tasks are unaffected by the requested confidence level (e.g., 30% confidence, 60% confidence or 90% confidence). Similarly, Fellner and Krügel (2012) found that various measurements of overprecision lead to different and unrelated results, concluding that "one cannot rely on the well-established miscalibration bias" (p. 142).

individual imagined the outcome of a specified random generating function a sufficiently large number of times, then an imagined set of outcomes would approximate the actual distribution. Most individuals are unlikely to mentally simulate an adequate number of outcomes due to cognitive limitations and bounded rationality. Thus, individuals are likely to generate a small set of mental samples when mentally visualizing realizations of a distribution. According to the "2-error" hypothesis, people's overgeneralization leads to overestimating rare events in one-shot decisions (the focus of this study). However, small samples can counteract this behavior by reducing the weight of rare outcomes (Marchiori et al., 2015). In addition, cognitive limitations and bounded rationality also imply that individuals often use a small sample of outcomes to represent a complete distribution (Juslin et al., 2007; Tong & Feiler, 2017).[2] Assuming that distribution can be estimated from encountered small samples, individuals may underestimate the true variance, leading to overprecise judgements (Juslin et al., 2007; Tong & Feiler, 2017). Therefore, overprecision is a natural result of misinterpreting the statistical properties of small samples.

Accordingly, I hypothesized that prompting individuals to mentally visualize the outcomes of a probability function will have greater estimates of values near the center of the distribution and smaller estimates of values in the distribution's tails relative to a control group. Thus, mental simulation can increase overprecision or decrease underprecision for distribution where estimates are typically underprecise. Also, visualizing an individual outcome can create a natural anchor due to saliency. As probabilities are often derived from an anchor (Einhorn & Hogarth, 1986), the anchor will increase the perceived likelihood of nearby outcomes. Consequently, a visualized outcome closer to the distribution center will likely increase overprecision. However, this may rely on people producing a minimal number of observations. While statistically, the average result is the most likely outcome, it does not align with typical expectations of a series of outcomes generated from a random device (Ayton & Fischer, 2004; Kahneman & Tversky, 1972). Thus, mentally simulating a larger number of outcomes could diminish precision, as participants will recall a more extensive array of outcomes and ascribe less weight to the middle of the estimated distributions.

I designed three preregistered experiments to test these predictions. In Study 1, participants estimated a binomial distribution characterized by 10 trials and a probability of 50% using a ball drop machine known as a quincunx. Participants were randomly assigned to a control or treatment condition, where they were asked to visualize an outcome from the ball drop machine. The results showed that mentally simulating the outcomes increased probability estimates around the mean, with probability estimates decreasing in the tails compared to the control group. Thus, participants engaged in greater visual processing decreased their perceived distribution variance (a hallmark of overprecision). Also, consistent with my hypotheses, I found evidence that visualizing an outcome close to the mean increases overprecision within the treatment group.

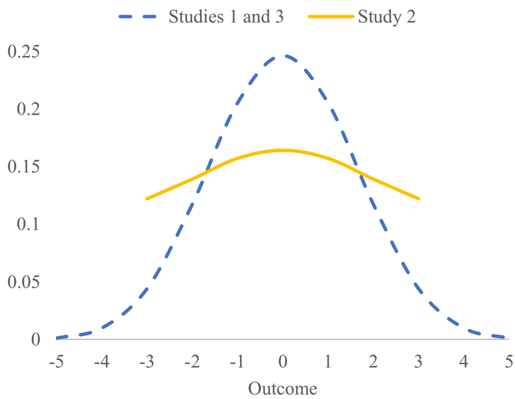
Although I concluded that mental simulation increases overprecision compared to the control group, like Moore et al. (2015), participants' binomial distribution estimates are "catastrophically" underprecise. The higher estimates around the mean imply that the visual simulation group's estimates are more accurate. Therefore, overprecision conflates with accuracy. In Study 2, I modified the ball drop machine to create a flatter distribution that maintains a central mode to disentangle overprecision from accuracy. Figure 1 shows the distributions of the two ball drop machines.

Consistent with the Study 1, the results of Study 2 showed that visual simulations lead to higher estimates of central values and smaller estimates of tail values relative to the control group. Moreover, participants in the control group were well-calibrated: Their central and tail estimates did not differ from the true values. In contrast, the visual group had overprecise estimates of central values and underestimated tail outcomes compared to the actual values. Again, I find participants who visualize an outcome closer to the mean are more overprecise.

A key element of the previous studies is that participants mentally simulate a single outcome, creating a clear anchor. However, if participants mentally simulate multiple outcomes, the representative heuristic may likely result in imagining draws farther from the mean. Study 3 tests this

[2] There are many benefits to using small samples related to bounded rationality, such as limiting memory (Barron & Erev, 2003 and simplifying tasks, Fiedler, 2000). Also, relying on small samples can be effective in dynamic environments (Plonsky et al., 2015).

Figure 1
Distributions of the Ball Drop Machines Outcomes for Study 1 and Study 2



Note. See the online article for the color version of this figure.

boundary condition by replicating Study 1 with one key differentiator—participants in the treatment group are asked to imagine five outcomes. While the perceived distribution of the treatment group resembles Study 1, the total weight in the middle and tails of the distribution does not differ from the control group. Moreover, the wider the variance of the imagined outcomes, the less precise the participants' estimated distributions.

My research contributes to the literature by identifying mental simulation of outcomes as a cognitive mechanism leading to differences in subjective probability estimates of aleatory distributions. This study also shows that previous findings of severe underprecision when estimating binomial distributions are robust. However, I show that Moore et al.'s (2015) findings on individuals' underprecision tendencies with estimating aleatory uncertainty are distribution-dependent. Participants can be well-calibrated when estimating other distributions, with a tendency toward overprecision (when primed with visual simulation).

Study 1

This study examines whether mental simulation of outcomes influences overprecision when estimating a distribution with aleatory uncertainty. For distributions, overprecision manifests in an underestimation of variance, which leads to higher expectations of mean outcomes (Ren & Croson, 2013). The study is a between-subject experiment where participants were randomly assigned to a control or a treatment group. In the control group, I elicit probabilities for all possible outcomes, known as SPIES (Haran et al., 2010). The treatment group introduces an intervention that explicitly prompts the participant to visualize the scenario before providing their explicit probability distribution. Using SPIES enables a comparison of the probabilities of the tails and center between control and treatment groups to determine if greater visual processing leads to greater overprecision. SPIES is also more effective than item confidence, which can suggest overprecision due to misspecified centered values while having underprecise distribution estimates (Moore et al., 2015). Note that the sole variation for the treatment group was the inclusion of mentally simulating a specific outcome.

Method

Participants

The target sample size is 100 per treatment group to ensure sufficient statistical power (>.90) to detect a medium-to-large-sized effect. Therefore, the planned sample size is 200 participants. In total, I recruited 200 American Mechanical Turk[3] participants (age: $M = 40.5$, $SD = 11.8$) with an incentive of $1, where 82 participants identified as Female.

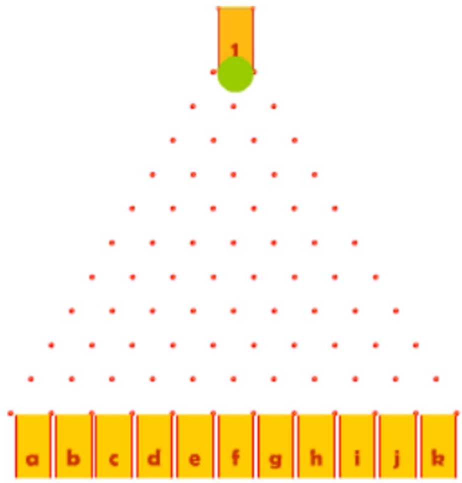
Procedure

Before starting the experiment, participants read an attention-check question.[4] Participants are then introduced to a ball drop machine known as a quincunx. In this experiment, the quincunx involves a ball bouncing down 10 rows of pegs, where it is equally likely to fall to the left or right at each juncture, eventually landing in one of 11 buckets (labeled A to K). The quincunx machine was also used in Studies 2–5 of Moore et al. (2015) and is presented in Figure 2.

[3] The participants for Studies 1 and 2 were screened to have a human intelligence task approval rate greater than 98% and to have completed at least 1,000 approved human intelligence tasks.

[4] The question asks "approximately how many cups of water do you drink each day" and is followed by a brief description about drinking water with the instructions "when answering the question below, you must select other and enter the phrase soda."

Figure 2
The Image of the Quincunx Machine Presented to Participants



Note. See the online article for the color version of this figure.

To elicit probability estimates for each outcome, participants are asked: "How likely is it that the ball will land in each of the different bins? In other words, if we dropped 100 balls, what percentage of the balls would land in each of the bins?" In the visual treatment group, participants are asked to "visualize a ball being dropped in the machine. Picture the ball moving either to the left or right at each peg so that it creates a path from the top to one of the bottom bins" before estimating the probabilities. Following Moore et al. (2015), I standardized participants' probabilities that exceeded 100% by dividing each estimate by the sum of the estimates. Thus, the probability estimates total 100% while maintaining the participants' original proportions.

To check the success of the intervention in promoting engagement in visual processing of outcomes, participants answered the following two questions (adapted from Yan et al., 2016): "While answering the previous questions, I imagined the ball dropping down the machine to a great extent?," and "When responding to the scenario, I experienced a lot of images of the ball dropping down the machine." The answers to both questions used a 7-point Likert scale ranging from *strongly disagree* to *strongly agree*. Finally, participants indicated their age and identified gender.

Probability Measures

I compared the tail and middle probabilities between the participant groups to test whether the visual group was more precise than the control group. For each experiment, participants estimate probabilities across $B + 1$ bins. Let P_i be the standardized probability of bin $i \in \{0,B\}$. (Note that the bins are relabeled to be numbered from 0 to B.) To measure tail probabilities, I considered $T_n = (\sum_{i=0}^{n-1} P_i + \sum_{i=B-n+1}^{B} P_i)/2n$, where I varied n depending on the number of bins B. Similarly, to measure probabilities near the distribution's center, I used $M_n = \sum_{i=B/2-n+1}^{B/2+n-1} P_i/(1 + 2(n - 1))$. Specifically, in Study 1, I compared $T_2 = \frac{1}{4}(p_0 + p_1 + p_9 + p_{10})$ and $M_3 = \frac{1}{5}(p_3 + p_4 + p_5 + p_6 + p_7)$. Generally, greater overprecision should decrease the probability of T and increase the probability of M.

Results

I averaged the values from the two manipulation check questions to check the manipulation. A comparison of the average responses between the control ($M = 5.25$, $SD = 1.51$) and the treatment group ($M = 5.78$, $SD = 1.11$), a difference of 0.53 (95% CI [0.16, 0.90]), suggests that the intervention successfully prompted the treatment group to engage in greater visual processing of outcomes, $t(198) = 2.82$, Cohen's $d = .40$, $p = .005$.

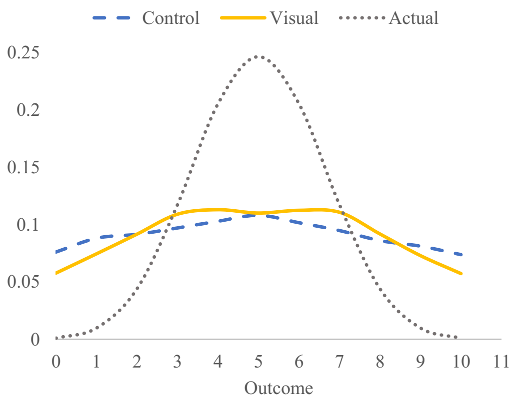
Figure 3 plots the average standardized[5] subjective probability estimates for the control and treatment groups. The distributions show that the control and treatment have similar estimates for the middle bin. Beyond the middle estimates, the distributions are quite different, with the treatment group underestimating outcomes in the tails of the distributions and overestimating outcomes in the middle of the distribution compared to the control group.

Table 1 provides the difference in probability estimates across the tail and central measures of overprecision. Consistent with the observations regarding Figure 3, there is a significant

[5] Recall that the probabilities are standardized. In Study 1, more than 50% of participants provided probabilities totaling 110 or less, and more than 75% provided probabilities that summed to 200 or less. In Study 2, over 65% of participants had total probabilities of 110 or less, and 83.5% reported total probabilities of less than 200.

Figure 3

Mean Probability Estimates for Each Bin in Study 1



Note. See the online article for the color version of this figure.

difference between the estimates around the tail. Similarly, the measures around the distribution's middle also differed between the control and treatment groups. (Supplemental Table SM.1 compares all other values of T_n and M_n, providing additional support for the current findings.[6]) Thus, I concluded that visual simulation increases participants' precision level compared to the control group.

The experiment asked participants in the treatment group to picture the ball dropping down the quincunx machine, explicitly asking them to identify the bin in which they visualize the ball landing. If the visual stimulation drove the precision of estimates, then the ball landing closer to the middle bin should correlate with greater overprecision. I examined this relationship by creating a variable called imagined outcome, which ranges from 0 to 5, where picturing the ball landing in the middle bin corresponds to a 5, picturing the ball landing one bin away from the middle corresponds to a 4, and so forth, where picturing the ball landing in one of the two end bins results in a 0. Thus, a higher value for the imagined outcome variable should lead to greater overprecision. (Recall that only participants in the treatment group answered the question on the imagined outcome.)

To test this, I conducted a simple regression of the relationship between the imagined outcome on M.[7] Table 2, which presents the regression results, shows that imagining the ball landing closer to the middle bin increased the average estimates near the distribution's center, indicating greater overprecision. Thus, imagining the ball landing closer to the middle bin correlates with greater overprecision. I also examined the number of participants providing flat distributions. In total, 22 participants provided a flat distribution (e.g., all 9s). Of these participants, 17 were in the control group, and five were in the visual group. This suggests that mental simulation could lead people to a more accurate distribution when participants are underprecise due to misunderstanding the probability-generating mechanism.

Study 2

The results of Study 1 suggest that mental simulation can increase estimates of central probabilities. However, the higher central estimates increase accuracy in estimating the distribution. Therefore, visualizing outcomes improve accuracy when estimating distributions is a plausible alternative interpretation of the results. To limit the degree of participants' underprecision, I adjusted the ball drop machine to produce a flatter distribution. If participants are not underprecise, higher estimates of central values due to mental simulation would represent worse accuracy and greater overprecision.

Method

Participants

Again, the target sample size is 100 per treatment group. In total, I recruited 200 American Mechanical Turk participants (88 participants identified as Female and three identified as nonbinary or prefer not to say; age: $M = 42.73$, $SD = 11.10$) with an incentive of \$1.

Procedure

The procedure (verbatim) followed Study 1, with three exceptions. First, I adjusted the ball drop mechanism to resemble a Plinko Machine.

[6] The results are largely independent of the metrics chosen for the tail distributions. However, for Study 1, there is no difference between the control and treatment group for middle values that have larger weight on the modal value (e.g., M_1 and M_2).

[7] I also conducted regressions for the other measures of overprecision for each study, available in the Supplemental Tables SM.2, SM.5, and SM.7.

Table 1
Average Estimates and Comparison of Probability Measures for All Studies

Metric	Study 1		Study 2		Study 3	
	Tail	Middle	Tail	Middle	Tail	Middle
Control						
M	0.080	0.101	0.120	0.171	0.074	0.105
SD	0.031	0.028	0.005	0.06	0.029	0.028
Treatment						
M	0.066	0.111	0.104	0.184	0.070	0.110
SD	0.028	0.027	0.005	0.072	0.033	0.030
Comparison						
Δ	−0.014***	0.010*	−0.016*	0.014*	−0.004	0.006
CI	[−0.022, −0.006]	[0.002, 0.018]	[−0.030, −0.002]	[0.002, 0.025]	[−0.013, 0.005]	[−0.003, 0.014]
t	−3.383	2.604	−2.247	2.335	−0.923	1.351

Note. CI = confidence intervals.
* $p < .05$. *** $p < .001$.

Rather than having one central release point for the ball to drop, I created six starting points. Figure 4 presents the image of the ball drop machine given to participants. Participants are told,

> The machine releases one ball randomly from 1 of the 6 spots above. It is equally likely to release the ball from any of the numbered spots. While the ball moves left or right with equal probability across most pegs, if the ball hits a peg adjacent to a wall, it can only move in the available direction.

Note that there are seven rows of pegs and seven bins (labeled A–G). Thus, for this study, $B = 6$, and the probability measures for comparison were $T_1 = \frac{1}{2}(p_0 + p_6)$ and $M_2 = \frac{1}{3}(p_2 + p_3 + p_4)$.

The second change to the experiment was the visual treatment. Before asking participants to visualize the ball bouncing down the pegs, participants imagined the ball was released from one of the starting points and then answered, "When you imagined the ball being dropped, from which position was the ball released?" Then, I asked them to imagine the ball dropping from the identified release point, visualizing the ball moving either to the left or right at each peg to create a path from the top to one of the bottom bins before asking which bin they visualized the ball landing in (which is the same question as in Study 1). The third change was an additional manipulation question stating, "When responding to the scenario, I engaged deeply in imagining the ball dropping down the machine," answered on a 7-point Likert scale.

Table 2
Regressions of the Mental Simulation on the Overprecision Measures

Metric	Study 1	Study 2	Study 3
Constant			
B	0.080***	0.131***	0.130***
SE	0.008	0.011	0.005
t	9.743	11.960	27.246
Mental simulation			
B	0.010***	0.024***	−0.003***
SE	0.002	0.005	0.001
t	3.947	4.330	−4.956

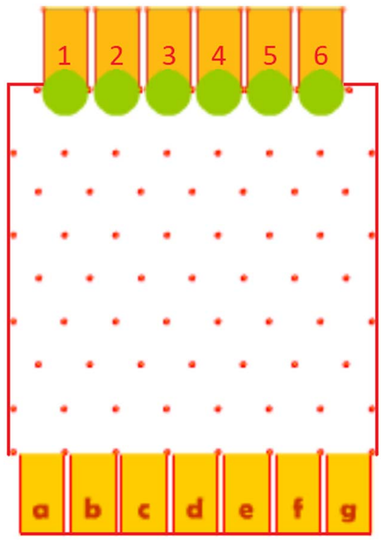
Note. For Studies 1 and 2, mental simulation pertains to how close the imagined outcome is to the center of the distribution. For Study 3, mental simulation pertains to the variance across the imagined outcome. *SE* = standard error.
*** $p < .001$.

Results

To check the manipulation, I averaged the values from the manipulation check questions. A comparison of the average responses between the control ($M = 5.05$, $SD = 1.63$) and the treatment group ($M = 6.07$, $SD = 1.03$), a difference of 1.02 (95% CI [0.64, 1.40]), suggests that the intervention successfully prompted the treatment group to engage in greater visual simulation, $t(198) = 5.26$, Cohen's $d = 0.76$, $p < .001$.

Figure 5 shows the average probability estimates for the Plinko machine. As in Study 1, the treatment group had lower estimates of the tail probabilities and higher estimates of the middle probabilities relative to the control group. Also, consistent with Study 1, Table 1 shows a significant difference between the control and

Figure 4
The Image of the Plinko Machine Presented to Participants

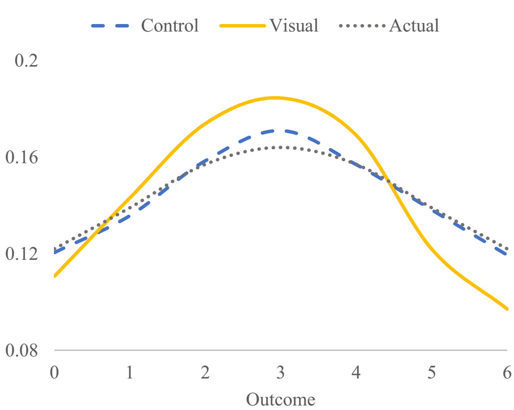


Note. See the online article for the color version of this figure.

treatment groups across the tail and central bin estimates.

A comparison with Figure 1 indicates that the control group was well-calibrated to the actual probabilities, while the treatment group seems overprecise. I compare the subjects' probability estimates across the four measures of overprecision against the true values using one-sample t tests. The results (see Supplemental Table SM.4) provide empirical support that the treatment group were overprecise, while the control group's estimates are not statistically different from the true values. Thus, contrasting Study 1, where participants were underprecise and the control group was more underprecise compared to the treatment group, in Study 2, the visual treatment group was overprecise compared to the distribution, in addition to having higher mean values than the control group. Thus, the visual treatment group is less accurate than the control group. Accordingly, I concluded that visual simulation led participants to increase their precision level compared to the control group.

Figure 5
Mean Probability Estimates for Each Bin in Study 2



Note. See the online article for the color version of this figure.

I also tested the relationship between the overprecision measures and the participants' imagined outcomes.[8] Following the analysis of Study 1, I conducted a simple regression for each measure of overprecision with the imagined outcome as the independent variable. Table 2 presents the regression models. Imagining the ball landing closer to the middle decreased the probability estimates for the tail probabilities and increased probability estimates near the center. Therefore, consistent with Study 1, imagining the ball landing closer to the middle bin correlates with greater overprecision, providing additional support that the visual simulation is responsible for driving the higher precision of estimates. I also examined the number of participants providing flat distributions. In total, 15 participants provided a flat distribution (e.g., all 14s). Of these participants, 12 were in the control group, and three were in the visual group. This suggests that mental stimulation can lead to less flat distributions by allocating more weight to imagined outcomes.

Study 3

This study tests whether the effect of mental simulation on overprecision holds when participants simulate several outcomes instead of one. Mentally picturing multiple outcomes could widen the range of imagined values, weakening the effect seen in prior studies, where focusing on a central outcome increased precision. This broader mental spread may lead to less

[8] The imagined outcome variable ranged from 0 to 3 due to the smaller range in Study 2. The ball landing in the middle bin corresponds to a 3, the ball landing one bin away from the middle corresponds to a 2, and so forth.

concentrated estimates and reduce the over-precision observed in earlier treatments.

Method

Participants

Again, the target sample size is 100 per treatment group. In total, I recruited 200 United Kingdom or American participants from Prolific (103 participants identified as female and two identified as nonbinary or prefer not to say; age: $M = 39.50$, $SD = 11.40$) with an incentive of 0.85 GBP. Three participants were removed due to excessively high estimates[9] for a single bin.

Procedure

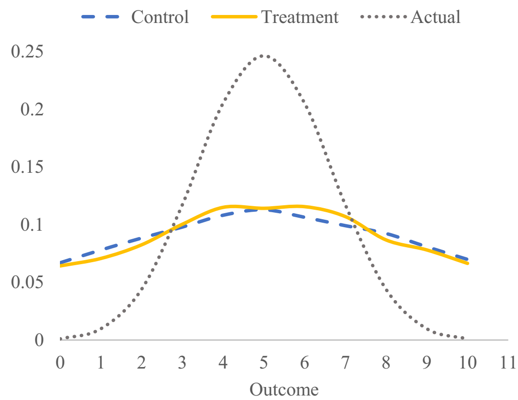
This experiment follows the procedure from Study 1, with one essential modification for the treatment group. Participants in the treatment group were asked to visualize the ball drop five times. After reading the ball drop machine instructions, they were informed, "You are going to do a short task 5 times in a row." They then were given the same instructions as in Study 1, but with a counter displayed (e.g., "1/5") to indicate which repetition they were completing.

Results

To assess the success of the intervention in promoting visual simulation, I compared the average responses between the control group ($M = 4.99$, $SD = 1.50$) and the treatment group ($M = 6.01$, $SD = 0.86$). A comparison of these means shows a significant difference of 1.02 (95% CI [0.67, 1.36]), indicating that the treatment group engaged more deeply in visual simulation. This difference was statistically significant, $t(195) = 5.78$, $p < .001$, with a large effect size, Cohen's $d = 0.82$.

Figure 6 plots the average standardized subjective probability estimates for the control and treatment groups in Study 3. Similar to Figure 3, the treatment and control groups show comparable estimates for the middle bin, with bins adjacent to the middle receiving higher estimates in the visual treatment group and lower estimates in the tails. However, the differences between the control and treatment groups are smaller than those observed in Study 1. The regression results in Table 1 further show

Figure 6
Mean Probability Estimates for Each Bin in Study 3



Note. See the online article for the color version of this figure.

that the differences between the control and treatment groups were not statistically significant. This suggests that multiple mental simulations are not as effective in generating the same level of precision as a single mental simulation.

The lack of a treatment effect may stem from participants imagining a wider range of outcomes when conducting multiple mental simulations. If participants interpret a greater variance in imagined outcomes as representative, despite these outcomes not being more statistically likely, then this wider mental distribution may reduce precision compared to imagining a single outcome. To test this, I conducted a regression using the average weight across the middle bins as the dependent variable. Unlike the previous studies, I considered the variance of the imagined outcomes as the independent variable. The results in Table 2 show that greater variance in imagined landing bins decreases the average estimates near the center of the distribution, supporting the rationale that increased imagined variance leads to a less concentrated mental representation, diminishing central precision. Again, I also examined the number of participants providing flat distributions. In total, 31 participants provided a flat distribution, with 21 in the control group and 10 in the treatment group. This is a larger proportion

[9] These participants all had estimates in one bin that exceeded 0.5 (in fact all three had estimates exceeding 2/3s in a single bin). For reference, no participant in Study 1 had an estimate that exceeded 0.5, and only two participants had estimates that exceeded 1/3. The results do not qualitatively change when including these participants.

for the treatment group than in Study 1, further supporting the finding that simulating multiple outcomes dilutes the precision compared to a single mental simulation.

Discussion

This article presents results from three preregistered experiments testing the impact of imagining probabilistic outcomes. In Study 1, I replicated Moore et al.'s (2015) findings that participants were "catastrophically" underprecise when estimating a binomial distribution, which also supports that individuals overgeneralize rare outcomes in one-shot events (Marchiori et al., 2015). For instance, the average probability estimates for the most extreme bins (i.e., T_1) are 7.5% and 5.7% in the control and visual treatment groups, respectively, while the actual probability is less than 0.1%. Similarly, the probability estimates for the middle bins (i.e., M_1) are 10.8% and 11.0% in the control and treatment groups, while the actual probability is almost 25%. These results support that the participants interpreted estimating probabilities with a high degree of ambiguity (often related to factors like ignorance) as ambiguity increases the uniformity of perceived likelihoods across potential outcomes (Einhorn & Hogarth, 1986). The substantial degree of underprecision leads Moore et al. (2015) to conclude against naive statistics (Juslin et al., 2007) and anchoring and insufficient adjustment (Tversky & Kahneman, 1974) as drivers influencing overprecision.

I explicitly explore the visual simulation of outcomes as a mechanism for causing individual differences in precision. Consistent with my hypothesis, visually simulating a single random outcome increased the focus on mean occurrences, irrespective of individuals being underprecise or well-calibrated. Thus, I contribute to the literature by demonstrating that visual processing is a cognitive mechanism that can drive differences in probability distribution estimates. In addition, I provide evidence that visualized outcomes provide an anchor, where (relative) overprecision increases with the visualized outcome being closer to the distribution's middle.

Interestingly, while Moore et al. (2015) argued against anchoring and insufficient adjustment (Tversky & Kahneman, 1974) and naive statistics (Juslin et al., 2007), I leverage these behavioral tendencies in combination with visual processing, which provides the theoretical rationale for my hypothesis. Therefore, I show that the underprecision results from Moore et al. (2015) are not mutually exclusive to the role of naïve statistics and anchoring and adjustment processes in explaining aspects of over and underprecision. Considering a nonbinomial distribution in Study 2, I show that participants are not always underprecise when estimating aleatory uncertainty. This is consistent with Crosetto et al. (2020), who show that individuals exhibit a central tendency bias when estimating a uniform distribution. Also, Study 3 showed that Study 1's results depend on the number of mental simulations. Even simulating five outcomes resulted in a flatter distribution with lower middle estimates.

My research complements recent efforts to uncover cognitive drivers of overprecision. For example, López-Pérez et al. (2021) provided evidence that simplified mental models of probability spaces can lead to insufficient Bayesian updating, generating overprecise outcomes. Similarly, Moore (2023) proposed that neglecting the ways one can be incorrect leads to overprecision. These explanations complement my findings on the role of visual processing since both are consistent with naïve statistics of mental samples (e.g., Tong & Feiler, 2017). In addition, visualizing outcomes offers a similar effect to the standard role of experience and feedback, whereby both experience and feedback typically decrease weights on rare events (Cohen et al., 2020; Marchiori et al., 2015).

This research also relates to a recent study by Landwehr et al. (2023), who tested the role of processing fluency on probability estimates of random future events that do not have an associated probability distribution. They find mixed evidence linking fluency and uncertainty estimates and call for research to understand why and when fluency influences future probability judgments. This study partly addresses this call by supporting that processing fluency related to visual processing outcomes can increase probability estimates of likely events under aleatory[10]

[10] In the Supplemental Materials, I conduct an additional experiment to examine whether the results can extend to epistemic uncertainty. The experiment supports the idea that visual simulations can also lead to narrower subjective probability distributions for epistemic uncertainty. Future research can also look at visual processing for distributions that combine aleatory and epistemic uncertainties (e.g., Soll et al., 2023).

events. Future research can formally integrate and unify these various explanations for generating differences in overprecision. In addition, future research can explore whether my results hold across other types of probabilistic distributions and how results change with learning (Sanchez & Dunning, 2018, 2020). For example, my results suggest that for beginners making decisions related to problems with aleatory uncertainty, visualizing outcomes could create a "beginner's bubble," raising distribution confidence (which can be beneficial for scenarios characterized by overgeneralization). Future research can incorporate visual processing into studies on metacognitive judgments to elicit the degree to which visual processing influences overprecision (Fleming & Lau, 2014) and to deepen our understanding of the neural correlates of overprecision (Rahnev, 2021).

Another limitation is that each study involved a small set of outcomes (11 in Studies 1 and three and seven in Study 2). While Study 3 establishes a boundary condition on Study 1, the number of imagined outcomes that diminish the mental visualization effect could depend on the number of potential outcomes in the distribution. For example, even a small number of mental simulations (e.g., 5–10) with a distribution that contains 100 possible outcomes could still result in greater precision. Examining the interactions between the number of mental simulations and the number of distribution outcomes provides another avenue for future research.

Another exciting area for future research involves exploring how individual differences in cognitive processing and personality traits influence overprecision. New research by Lawson et al. (2023) represents a significant advancement in understanding how cognitive measures (e.g., cognitive reflection, numeracy, and open-minded thinking) and personality traits (e.g., dark triad) impact overprecision. Future work should examine how these individual-specific factors relate to visual processing and precision in probability judgments, providing a deeper understanding of the mechanisms driving overprecision.

References

Anderson, C. A. (1983). Imagination and expectation: The effect of imagining behavioral scripts on personal influences. *Journal of Personality and Social Psychology*, *45*(2), 293–305. https://doi.org/10.1037/0022-3514.45.2.293

Attari, A., Chatterjee, P., & Cabano, F. G. (2022). It happens because I'm watching it: Observing an uncertain event can affect subjective probability judgments. *Journal of Behavioral Decision Making*, *35*(3), Article e2269. https://doi.org/10.1002/bdm.2269

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition*, *32*(8), 1369–1378. https://doi.org/10.3758/BF03206327

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*(3), 215–233. https://doi.org/10.1002/bdm.443

Bilgin, B. (2012). Losses loom more likely than gains: Propensity to imagine losses increases their subjective probability. *Organizational Behavior and Human Decision Processes*, *118*(2), 203–215. https://doi.org/10.1016/j.obhdp.2012.03.008

Carroll, J. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, *14*(1), 88–96. https://doi.org/10.1016/0022-1031(78)90062-8

Cohen, D., Plonsky, O., & Erev, I. (2020). On the impact of experience on probability weighting in decisions under risk. *Decision*, *7*(2), 153–162. https://doi.org/10.1037/dec0000118

Crosetto, P., Filippin, A., Katuščák, P., & Smith, J. (2020). Central tendency bias in belief elicitation. *Journal of Economic Psychology*, *78*, Article 102273. https://doi.org/10.1016/j.joep.2020.102273

Einhorn, H. J., & Hogarth, R. M. (1986). Decision making under ambiguity. *The Journal of Business*, *59*(4), S225–S250. https://doi.org/10.1086/296364

Fellner, G., & Krügel, S. (2012). Judgmental overconfidence: Three measures, one bias? *Journal of Economic Psychology*, *33*(1), 142–154. https://doi.org/10.1016/j.joep.2011.07.008

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659–676. https://doi.org/10.1037/0033-295X.107.4.659

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, Article 443. https://doi.org/10.3389/fnhum.2014.00443

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology*, *43*(1), 89–99. https://doi.org/10.1037/0022-3514.43.1.89

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*(7), 467–476. https://doi.org/10.1017/S1930297500001637

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*(350), 271–289. https://doi.org/10.1080/01621459.1975.10479858

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, *114*(3), 678–703. https://doi.org/10.1037/0033-295X.114.3.678

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Landwehr, J. R., Winkler, S., & Bornemann, T. (2023). Feeling the future? Mixed empirical evidence for a link between processing fluency and judgments about future probability. *Social Psychological and Personality Science*, *14*(4), 428–436. https://doi.org/10.1177/19485506221105375

Langnickel, F., & Zeisberger, S. (2016). Do we measure overconfidence? A closer look at the interval production task. *Journal of Economic Behavior & Organization*, *128*, 121–133. https://doi.org/10.1016/j.jebo.2016.04.019

Lawson, A., Larrick, R. P., & Soll, J. B. (2023). *Forms of overconfidence: Reconciling divergent levels with consistent individual differences*. SSRN. https://doi.org/10.2139/ssrn.4558486

López-Pérez, R., Rodriguez-Moral, A., & Vorsatz, M. (2021). Simplified mental representations as a cause of overprecision. *Journal of Behavioral and Experimental Economics*, *92*, Article 101681. https://doi.org/10.1016/j.socec.2021.101681

Marchiori, D., Di Guida, S., & Erev, I. (2015). Noisy retrieval models of over-and undersensitivity to rare events. *Decision*, *2*(2), 82–106. https://doi.org/10.1037/dec0000023

Moore, D. A. (2023). Overprecision is a property of thinking systems. *Psychological Review*, *130*(5), 1339–1350. https://doi.org/10.1037/rev0000370

Moore, D. A., Carter, A. B., & Yang, H. H. (2015). Wide of the mark: Evidence on the underlying causes of overprecision in judgment. *Organizational Behavior and Human Decision Processes*, *131*, 110–120. https://doi.org/10.1016/j.obhdp.2015.09.003

Moore, D. A., & Dev, A. S. (2020). Overconfidence. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 3382–3386). Springer.

Plonsky, O., Teodorescu, K., & Erev, I. (2015). Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychological Review*, *122*(4), 621–647. https://doi.org/10.1037/a0039413

Rahnev, D. (2021). Visual metacognition: Measures, models, and neural correlates. *American Psychologist*, *76*(9), 1445–1453. https://doi.org/10.1037/amp0000937

Ren, Y., & Croson, R. (2013). Overconfidence in newsvendor orders: An experimental study. *Management Science*, *59*(11), 2502–2517. https://doi.org/10.1287/mnsc.2013.1715

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, *114*(1), 10–28. https://doi.org/10.1037/pspa0000102

Sanchez, C., & Dunning, D. (2020). Decision fluency and overconfidence among beginners. *Decision*, *7*(3), 225–237. https://doi.org/10.1037/dec0000122

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin*, *11*(1), 118–127. https://doi.org/10.1177/0146167285111011

Sherman, S. J., Skov, R. B., Hervitz, E. F., & Stock, C. B. (1981). The effects of explaining hypothetical future events: From possibility to probability to actuality and beyond. *Journal of Experimental Social Psychology*, *17*(2), 142–158. https://doi.org/10.1016/0022-1031(81)90011-1

Soll, J. B., Palley, A. B., Klayman, J., & Moore, D. A. (2023). Overconfidence in probability distributions: People know they don't know, but they don't know what to do about it. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2019.00660

Tong, J., & Feiler, D. (2017). A behavioral model of forecasting: Naive statistics on mental samples. *Management Science*, *63*(11), 3609–3627. https://doi.org/10.1287/mnsc.2016.2537

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Yan, D., Sengupta, J., & Hong, J. (2016). Why does psychological distance influence construal level? The role of processing mode. *Journal of Consumer Research*, *43*(4), 598–613. https://doi.org/10.1093/jcr/ucw045

Received May 20, 2023
Revision received September 11, 2024
Accepted September 12, 2024 ■