



Cognitive Science 46 (2022) e13144

© 2022 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13144

# Sensitivity to Evidential Dependencies in Judgments Under Uncertainty

Belinda Xie, Brett Hayes

*School of Psychology, University of New South Wales Sydney*

Received 8 September 2021; received in revised form 25 February 2022; accepted 6 April 2022

---

## Abstract

According to Bayesian models of judgment, testimony from independent informants has more evidential value than dependent testimony. Three experiments investigated learners' sensitivity to this distinction. Each experiment used a social version of the balls-and-urns task, in which participants judged which of two urns was the most likely source of evidence presented by multiple informants. Informants either provided independent testimony based solely on their own observations or dependent-sequential testimony that considered the testimonies of previous informants. Although participants updated their beliefs with additional evidence, this updating was generally insensitive to evidential dependency (Experiments 1 and 2). A notable exception was when individuals were separated according to their beliefs about the relative value of independent and sequential evidence. Those who viewed independent evidence as having greater value subsequently gave more weight to independent testimony in the balls-and-urns task (Experiment 3), in line with the predictions of a Bayesian model. Our findings suggest that only a minority of individuals conform to Bayesian predictions in the relative weighting of independent and dependent evidence in judgments under uncertainty.

*Keywords:* Social learning; Judgment; Bayesian modeling; Information dependence

---

---

**Funding Information:** This work was supported by the Australian Research Council Discovery Grant [DP190101224] awarded to Brett Hayes. Belinda Xie was supported by an Australian Government Research Training Program Scholarship and the UNSW Scientia PhD Scholarship Scheme.

Correspondence should be sent to Belinda Xie, School of Psychology, University of New South Wales Sydney, Sydney, NSW 2052, Australia. E-mail: belinda@xie1.net

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

## 1. Introduction

We cannot personally collect and interpret evidence for every decision we need to make, so we often rely on information from other people. Information shared by others can influence mental and physical health (Thoits, 2011), voter turnout (Bond et al., 2012), and risk perceptions of COVID-19 (Dryhurst et al., 2020). These are examples of social learning: reasoning based on information (or “testimony”) given by other people (“informants”) rather than reasoning based on directly observed evidence. Social learners often learn quicker or with greater accuracy than learners who do not receive social information (Derex & Boyd, 2018; Rendell et al., 2010; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012). However, the relative advantage of social learning depends on the nature of the learning environment and the behavior of informants (Perreault, Moya, & Boyd, 2012; Rendell et al., 2010). Social information must therefore be used selectively, in proportion to the evidential value of that social testimony.

Informants who are dependent typically provide less evidential value than informants who are independent (e.g., Hahn, von Sydow, & Merdes, 2019; Whalen, Griffiths, & Buchsbaum, 2018; Soll, 1999; Mercier & Miton, 2019; Lorenz, Rauhut, Schweitzer, & Helbing, 2011). Informants can be dependent if they use a common evidence source, share cognitive biases, or communicate with one other before providing testimony. For example, three friends who retweet the one original tweet (e.g., from *The New York Times*) are dependent on a shared evidence source. Compare this to three friends who each retweet a different tweet (e.g., from *The New York Times*, *The Washington Post*, and *The Wall Street Journal*). In most circumstances, you gain more information from reading the second set of independent retweets than you do from reading the first. The first set involves redundant repetitions, while the second set draws from a larger pool of evidence and is therefore more likely to provide a more accurate or representative sample of the underlying truth (Dietrich & Spiekermann, 2013; Hahn et al., 2019; Lorenz et al., 2011). As detailed below, the superior value of independent over dependent social evidence is often assumed by normative models of judgment and decision-making (e.g., Whalen et al., 2018; Anderson & Holt, 1997). The key issues addressed in the current work are the conditions under which people are sensitive to the epistemic warrant of independent information, compared to subtle forms of information dependency, and whether individuals differ in such sensitivity.

### 1.1. Reasoning with dependent information

Previous work has shown that people generally give less weight to dependent social evidence compared to independent evidence, when such evidence involves high levels of repetition or redundancy (Madsen, Hahn, & Pilditch, 2019; Mercier & Morin, 2019; Xie, Navarro, & Hayes, 2020). For example, Mercier and Miton (2019) showed that people were unlikely to be swayed by additional recounts of the same event (e.g., hearing additional retellings of an individual’s one mistake at work did not affect the perceived competency of that individual).

It is less clear, however, whether people attach different weights to independent evidence as compared to dependent evidence that contains a mixture of old and new information

(cf. Unkelbach, Fiedler, & Freytag, 2007; Yousif, Aboody, & Keil, 2019). An example of the latter would be when an individual repeats or “retweets” a message from another person, but adds to or embellishes that message. Likewise, reports from different newspapers that draw from the same primary source may add different details to their reports.

Such cases can be viewed as examples of sequential testimony. Sequential testimony occurs when multiple informants each have their own private evidence and access to the public testimonies of previous informants. Considerable work has examined judgments based on sequential testimony in tasks involving intuitive probability judgments. This work has often used a social variant of the classic “balls-and-urns” task introduced by Edwards, Peterson, and colleagues (Phillips, Hays, & Edwards, 1966; Peterson, Schneider, & Miller, 1965). In this task, learners use evidence provided across a sequence of trials to judge which of two uncertain states of the world (i.e., which of two urns with different proportions of red and blue balls has been selected in a random draw) is more likely to be true. On each trial, the learner receives evidence in the form of a ball sampled from the source urn (e.g., a red ball). In the social variant of the task (Anderson & Holt, 1997; Whalen et al., 2018), the evidence on each trial comes from an informant who has drawn a ball (whose color is not revealed) and who makes a judgment about the most likely urn (e.g., “I looked at my ball and I guess the red urn was selected”). Sequential testimony is induced by allowing each informant in the sequence to hear the judgments of previous informants before making their own judgment.

Previous studies have shown that sequential testimony can trigger information cascades, with those making judgments later in the testimony chain copying the majority view conveyed by previous informants, rather than basing their judgments on their own private evidence (Banerjee, 1992). Such cascades are thought to contribute to social fads that have inflated the popularity of political candidates, financial forecasts, and scientific theories (Bikhchandani, Hirshleifer, & Welch, 1992; Bloomfield & Hales, 2009).

What remains unclear, however, is whether people discriminate between the evidential value of such sequential testimony as compared to an equivalent amount of independent testimony (e.g., where each informant provides a judgment independently of other informants). As detailed below, normative Bayesian models of judgment prescribe that, in most circumstances, people should give more weight to independent testimony than sequentially dependent testimony. Previous empirical work on this issue, however, has provided equivocal results about whether people are sensitive to this distinction.

## 1.2. *Independent versus dependent sequential testimony*

Whalen et al. (2018) outlined a Bayesian model of judgment using a sequential balls-and-urns task. In this task, learners receive some directly observed data  $d$  (e.g., a ball drawn from an urn) and social testimony from  $n$  informants ( $t_1, \dots, t_n$ ). In this model, the posterior probability of each hypothesis about the source urn (e.g., red urn or blue urn) is proportional to (1) the prior probability of this hypothesis,  $p(h)$ , (2) the probability of receiving the social testimonies if this hypothesis were true,  $p(t_1, \dots, t_n | d, h)$  and (3) the probability of receiving the directly observed ball if this hypothesis were true,  $p(d | h)$ . In the dependent sequential

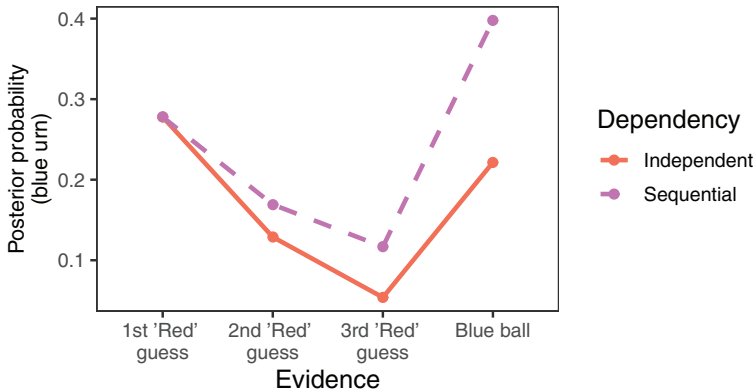


Fig. 1. The posterior probability that the selected urn was the blue urn with independent or sequential testimony, from Whalen et al.’s (2018) model of the judgment under uncertainty. As evidence accumulates along the x-axis, the probability that the selected urn is the blue urn decreases with additional “red” testimonies, but increases again with a final blue “own-ball.” Independent testimonies produce a greater decrease in probabilities than sequential testimonies, and a smaller increase following a contradictory own-ball because independent testimonies carry more evidential weight than sequential testimonies.

condition, each informant’s testimony depends on any preceding testimonies  $t_i$  depends on any preceding testimonies  $t_1, \dots, t_{i-1}$ . Thus, although the first informant interprets their own ball based on equal priors for both hypotheses (because the urn was randomly selected), subsequent informants will interpret their own ball based on revised priors that account for previous testimonies (see Eq. 1). This causes informants later in the sequence to become relatively more influenced by previous testimonies and less influenced by their own ball.

$$p(t_1, \dots, t_n | h) = p(t_1 | h) \prod_{i=2}^n p(t_i | t_1, \dots, t_{i-1}, h). \tag{1}$$

In contrast, when the informants are independent, the probability of an informant’s testimony is determined only by their private data ( $d_i$ ), and how they map these data to their testimony,<sup>1</sup> as shown in Eq. 2.

$$p(t_i | h) = \sum_{d_i} p(d_i | h)p(t_i | d_i). \tag{2}$$

To illustrate the contrasting predictions for sequential and independent testimonies, these equations were applied to source urns with complementary 5/6 and 1/6 distributions of different colored balls. On each of three testimony trials, the informant, after receiving their own private evidence drawn from the source urn, judges one of the urns (e.g., red) to be the most likely source. There is also a final “own ball” trial, where the participant, who previously received the testimonies, draws a different colored ball (e.g., a blue ball).

The resulting posterior probabilities for each testimony trial are given in Fig. 1 (see <https://osf.io/qng93/> for derivations). As each informant judges in favor of the red urn, the

posterior probability of the blue urn being the source decreases. Note, however, that this decrease is steeper across successive trials for independent than sequential testimony. When a ball of the contrasting blue color is received by the participant, there is a larger shift in favor of the blue urn in the sequential than the independent condition.

Whalen et al. (2018) tested these predictions in an experiment where participants received either three sequential or three independent testimonies. Whalen et al. (2018) also included a “shared” testimony condition, where all informants made judgments based on a single ball that was shared between them. Informant testimonies always supported the same urn (e.g., the red urn as the source urn), before participants received their own ball of the opposite color (a blue ball). After receiving this evidence, participants rated the relative likelihood that the balls had been drawn from the red or blue urn.

Likelihood ratings on the final trial indicated that participants attached more weight to independent than shared evidence. However, likelihood ratings did not differ between independent and sequential conditions (based on a frequentist statistical test) in two of the three experiments conducted by Whalen et al. (2018). This suggests that, contrary to the Bayesian account, people do not readily distinguish between the evidential value of independent and sequential testimony.

One exception to this trend occurred only after a substantial change to the experimental paradigm. Participants’ ratings did differ in the independent and sequential conditions when: i) the evidence sample included a “diagnostic” ball that revealed the true urn with certainty (i.e., a uniquely colored ball that was only found in a single urn), and ii) testimony was not corroborating (the final informant always dissented from the previous two, presumably because they drew the diagnostic ball). This result is consistent with other studies using dissenting testimonies that have shown that learners can (imperfectly) distinguish between independent and dependent testimonies when they conflict with each other (e.g., Otsubo, Whalen, & Buchsbaum, 2017; Pilditch, Hahn, Fenton, & Lagnado, 2020). In our view, however, this Whalen et al. (2018) experiment addresses a fundamentally different issue from their previous studies comparing sequential and independent evidence. As soon as the sequential testimony suggested the presence of the diagnostic ball, participants could discard evidence from any previous testimony. In other words, the experiment primarily tested whether participants were sensitive to the implications of perfectly predictive cues (the diagnostic ball) over probabilistic evidence. Hence, thus far there is no clear evidence that learners distinguish between the evidential value of independent and sequential testimonies.

### *1.3. The current work*

The aim of the current work was to advance our understanding of how people use independent testimony compared to sequential testimony in judgments under uncertainty, and whether all individuals do so in the same way. To address these questions, we extended the paradigm used by Whalen et al. (2018) in a number of ways. First, whereas Whalen et al. (2018) only measured intuitive judgments of probability at the end of the evidence chain (i.e., after the participant had heard informant testimonies and drew their own ball), we

queried participants' probability judgments on each evidence trial. This richer data set allowed for a more sensitive test of the contrasting Bayesian predictions in the sequential and independent testimony conditions. For example, we could examine whether conditions differed in the downward revision of probability estimates during the testimony stage (cf. left side of Fig. 1) and in their "uptick" in probability estimates following the own-ball (cf. the right side of Fig. 1).

The second advantage of more fine-grained comparisons between participant judgments and model predictions is that this can reveal *why* people may fail to differentiate between sequential and independent evidence. Such insensitivity may mean that people fail to discount the information dependence of sequential testimony, treating sequential testimony in a similar way to independent testimony. This echoes arguments advanced by researchers examining information dependency in other tasks, who suggest that people tend to reason appropriately with the more straightforward case of independent evidence, but struggle to make the necessary corrections for dependency (e.g., Enke & Zimmermann, 2017; Unkelbach & Rom, 2017; Yousif et al., 2019).

On the other hand, it is possible that people do not sufficiently revise their probability estimates when presented with independent evidence. Such conservatism in belief updating has often been observed in non-social probability judgment tasks (Navon, 1978; Phillips & Edwards, 1966). In the current paradigm, conservatism would mean that the observed estimates in the independent condition resemble those predicted for the sequential condition. These two alternatives—failing to discount dependent evidence or being conservative with independent evidence—could not be easily distinguished in the one-off ratings obtained by Whalen et al. (2018). In the current studies, however, we can distinguish them by comparing the relative fit of Bayesian model predictions to the data from the sequential and independent conditions, respectively.

A third important innovation in the current work was that we added elements to the design that allowed for stronger tests of the Bayesian model predictions. In Experiment 1 and subsequent studies, we added informants to the evidence chain, amplifying the predicted differences between probability estimates in the sequential and independent conditions. In Experiment 2, we also added incentives for accurate performance and used a logarithmic scale for intuitive probability estimates.

Finally, in Experiment 3, we examined the possibility of individual differences in sensitivity to the implications of sequential and independent evidence. This possibility was never considered by Whalen et al. (2018), but is suggested by other work. For example, individuals often differ in the extent to which they follow Bayesian prescriptions about the effects of sampling on property inference (e.g., Hayes, Navarro, Stephens, Ransom, & Dilevski, 2019; Navarro, Dry, & Lee, 2012) and classification (Ransom, Hendrickson, Perfors, & Navarro, 2018), and in how they combine prior probabilities and likelihoods when estimating probabilities (e.g., Brase, 2021; Sirota, Juanchich, & Hagemayer, 2014). This suggests the possibility that only a subset of individuals will show more belief updating in response to independent than sequential evidence.

## 2. Experiment 1

Experiment 1 reexamined people's sensitivity to the difference in the epistemic warrant of independent and sequential testimony. The study employed the balls-and-urns task used by Whalen et al. (2018), but with some notable modifications. First, we tracked participant judgments after each piece of evidence was presented. Second, we introduced conditions with five and seven corroborating informants, in addition to the original three-informant condition used by Whalen et al. (2018).

We derived predictions from the Bayesian model of judgment for each experimental condition and compared participant data against these predictions. This allowed us to examine the extent to which people in the sequential condition factored information dependence into their judgments, and whether those in the independent condition updated their beliefs in line with Bayesian prescriptions.

Bayesian statistical analyses were used in this and all subsequent studies (Wagenmakers et al., 2018). Bayesian analyses are particularly useful given that previous comparisons between likelihood ratings in sequential and independent evidence conditions found null results (Whalen et al., 2018). We used Bayesian analyses to produce Bayes factors that quantify the odds of the data occurring under the alternative or null hypotheses. The larger the Bayes factor ( $BF_{10}$ ), the greater the support for the alternative hypothesis. Conversely, the greater the inverse  $BF_{01}$ , the greater the support for the null hypothesis. As a guide, a  $BF_{10}$  from 1 to 3 provides inconclusive evidence for the alternative hypothesis compared to the null hypothesis, from 3–20 provides positive evidence, 30–150 provides strong evidence, and greater than 150 provides very strong evidence for the alternative hypothesis (Raftery, 1995).

### 2.1. Method

#### 2.1.1. Participants

Participants were recruited from MTurk ( $n = 216$ ) and Prolific ( $n = 120$ ). The MTurk sample consisted of U.S. residents, while the Prolific sample comprised residents of multiple countries. We found no evidence of qualitatively different patterns between the MTurk and Prolific samples, which is consistent with previous comparisons between the two platforms (Peer, Brandimarte, Samat, & Acquisti, 2017). The total sample size was 336 (mean age = 32.02,  $n = 122$  women,  $n = 212$  men,  $n = 2$  nonbinary). In Experiments 1 and 2, we aimed to achieve cell sample sizes that matched or exceeded those of Whalen et al. (2018) Experiment 1 (i.e., cell  $n \geq 50$ ). A further 195 U.S. residents were recruited from MTurk for a follow-up check of our independent/sequential manipulation (see Section 2.2.3 for details).

#### 2.1.2. Design and procedure

The study used a 2 (dependency: independent, sequential)  $\times$  3 (number of informants: three, five, seven) between-subjects design for a total of six conditions. The dependent variable was the mean likelihood ratings about the source urn on each evidence trial. Likelihood

ratings were scaled such that higher ratings indicated greater belief that the source urn was the urn implied by the participant's own-ball, while lower likelihood ratings indicated greater belief that the source urn was the urn implied by the informant testimonies. The experiment was administered online using jsPsych (v.6; de Leeuw, 2015). Code and materials for all experiments are available at <https://osf.io/qng93/>.

On-screen instructions first explained that the participant's goal was to use information provided by informants ("friends") and their own draw from the urn to judge which of two urns was more likely to be the source of all draws. Each urn was filled with 100 majority-colored balls (e.g., red) and 20 minority-colored balls (e.g., blue).

In the initial informant stage (Fig. 2A), participants were introduced to three, five, or seven friends. These friends were represented by silhouettes displayed below a name (selected randomly from the pool of Sarah, Mary, Ann, Emma, Jane, Claire, and Lucy). Participants were told that the friends also understood the ball proportions in the two urns and would help the participant by sharing their best guesses (testimonies) as to which urn was selected.

Informants were shown one ball from the selected urn, sampled with replacement. The participant never saw the color of these balls—they were represented visually by a gray circle (see Fig. 2A). In the independent condition, informants were each situated behind separate doors and each testimony was given after participants clicked on a door. In the sequential condition, informants were viewed together but gave testimony in turn. Example testimonies are shown in Fig. 2A. The testimonies always supported the same urn (e.g., each informant's judgment supported the "red jar").

After each testimony was received, participants provided a likelihood rating on an 11-point scale (Fig. 2B): "Based on my friends' guess(es), how likely is it that the selected jar was the blue jar or the red jar?" (scale: 0 = "Definitely the [jar implied by the testimonies]," 5 = "Equally likely to be the red or blue jar," and 10 = "Definitely the [jar implied by the participant's own-ball]"). After making each likelihood rating, participants received the next piece of evidence (informant testimony or own-ball).









After all social testimonies had been given, one additional ball was "randomly" sampled from the selected urn and shown directly to the participant (e.g., Fig. 2A). In fact, this "own-ball" was always the opposite color to the color implicated in informant testimonies. Again, after the own-ball on the final trial, participants made a likelihood rating. The total number of ratings made (four, six, or eight rating trials) varied with the number of informants (three, five, or seven). The experiment lasted around 5 min.

## 2.2. Results

We derived model predictions using the Bayesian model specified by Whalen et al. (2018), shown in Fig. 3A. The pattern of model predictions is similar to the more general description given in Fig. 1, but the predicted differences between likelihood ratings based on independent or sequential testimony are larger for the five- and seven-informant conditions than for the three-informant condition. The full code to generate these predictions can be found at <https://osf.io/qng93/>.



(A) Informant Testimonies & Own-Balls

		Ratings Trial			
		1	2	3	4
Independent	Ann	 "I looked at my ball and I guess the red jar was selected." I have not talked to the others, nor did I see any of the balls given to the others."	Jane  "I looked at my ball and I guess the red jar was selected." I have not talked to the others, nor did I see any of the balls given to the others."	Mary  "I looked at my ball and I guess the red jar was selected." I have not talked to the others, nor did I see any of the balls given to the others."	Here is your ball. 
	Sequential	Ann  "I looked at my ball and I guess the red jar was selected." I did not see any of the balls given to the others."	Jane  "I looked at my ball and I thought about what Ann said. I guess the red jar was selected." I did not see any of the balls given to the others."	Mary  "I looked at my ball and I thought about what Ann and Jane said. I guess the red jar was selected." I did not see any of the balls given to the others."	Here is your ball. 

(B) Likelihood Ratings

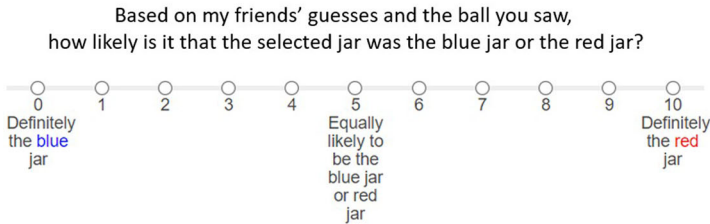


Fig. 2. Summary of Experiment 1 procedure. (A) Participants received informant testimony given in separate rooms (Independent) or in turn (Sequential). (B) Participants judged which urn was most likely to be the source of the draws. The example shown here is the final rating taken after all informant testimonies and the participant's own private ball was presented. A similar rating scale, omitting reference to the "the ball you saw," was presented after each testimony.

2.2.1. Likelihood ratings

For the Bayesian *t*-tests, ANOVAs and linear regressions reported here, we used the default Cauchy-distributed priors<sup>2</sup> of the BayesFactor R package (Morey & Rouder, 2018). Mean likelihood ratings for each rating trial are shown in Fig. 3B.

First, we assessed whether participants updated their beliefs about the most likely source urn as they received successive pieces of evidence. The slopes in Fig. 3B suggest that mean

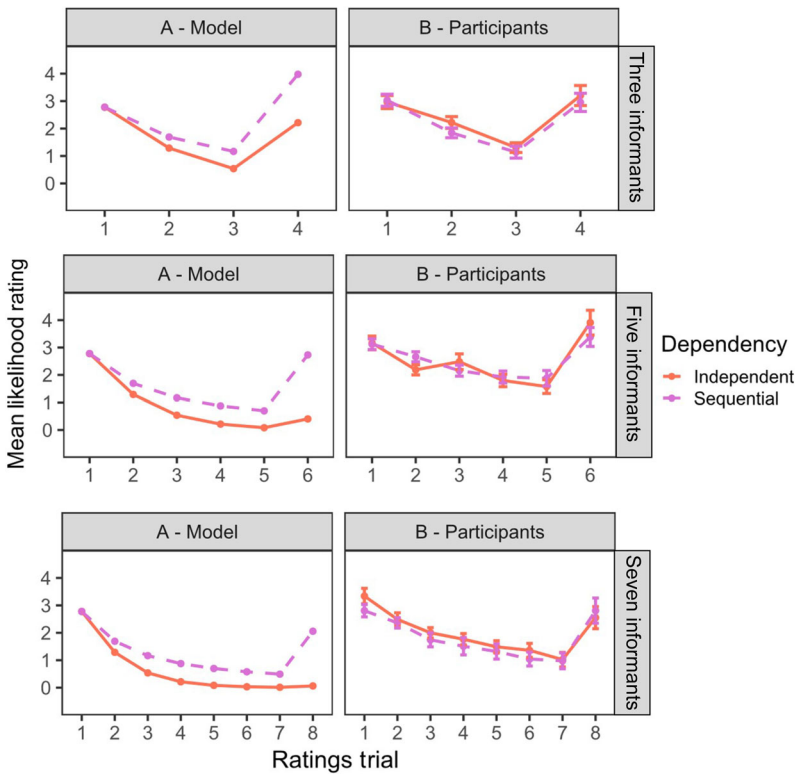


Fig. 3. Experiment 1 model predictions (A) and participants' mean likelihood ratings about the source urn (B) model predictions and mean ratings of the likelihood that the source urn matched the participant's private evidence after each piece of evidence was received, by the dependency condition and number of informants condition. In this and subsequent figures, model predictions show posterior probabilities multiplied by 10 and error bars represent  $\pm 1$  standard error.

likelihood ratings did indeed decrease during the majority testimony stage and then increase in the final trial, as per model predictions. In the majority testimony stage, a Bayesian linear model with rating trial as a predictor provided a better fit to participant data than models without this predictor, for each number-of-informants condition (all  $BF_{10} > 1000$ ). For the final uptick, Bayesian  $t$ -tests provided very strong evidence for an increase in ratings between the penultimate and final trials for each number-of-informants condition (all  $BF_{10} > 1000$ ).

The Bayesian model also predicted greater belief revision in the majority testimony stage as the number of informants increased. To examine this, we calculated the linear slope difference between ratings on the first and last testimony trials and examined whether these slopes differed with the number of informants. Contrary to predictions, there was inconclusive evidence for an effect of informant number ( $BF_{01} = 2.65$ ).

The most crucial issue concerns the effects of evidential dependency on belief updating. Fig. 3B suggests that the belief updating functions did not differ substantially between the two dependency conditions. During the majority testimony stage, likelihood ratings did not differ

between independent and sequential conditions in overall value (no main effect of dependency;  $BF_{01} = 5.94, 8.01, 2.25$  for three-, five-, and seven-informant conditions, respectively) nor steepness (no interaction between dependency and ratings trial;  $BF_{01} = 10.85, 30.59, 237$  for three-, five-, and seven-informant conditions). The uptick in the final trial also did not differ between independent and sequential conditions ( $BF_{01} = 5.21, 1.19, 4.40$  for three-, five-, and seven-informant conditions). Thus, participants' likelihood judgments were influenced by independent and sequential testimonies to a similar extent.

### 2.2.2. Comparing participant ratings with model predictions

Comparing the model predictions with observations in each row of Fig. 3 suggests that likelihood ratings in the sequential conditions were closest to model predictions. In contrast, ratings in the independent conditions were conservative—ratings did not decrease to the extent predicted by the model in the testimony stage. In the final trial, ratings in the independent conditions also showed a larger-than-predicted increase.

This conclusion was supported by a comparison of the Whalen et al. (2018) model predictions to the data in each dependency condition (note the model does not have any free parameters, so no parameters had to be estimated from the data). We used  $R^2$  as an overall measure of model fit.  $R^2$  values were derived from linear regressions in which the model prediction for each ratings trial was used to predict the mean likelihood rating for that trial. Two separate linear models were used for the two dependency conditions. As there were four, six, and eight ratings trials for three-, five-, and seven-informant conditions, there were 18 data points in each of the two linear models.

Participant ratings were closer to model predictions in the sequential conditions ( $R^2 = .74$ ) than the independent ( $R^2 = .44$ ) conditions. This difference in model fit was confirmed by a Bayesian linear regression where the Whalen et al. (2018) model predictions and dependency condition were entered as separate predictors of mean likelihood ratings. A regression model containing the dependency condition gave a better account of the data than a model without this predictor ( $BF_{10} = 23.45$ ).

### 2.2.3. Follow-up manipulation check

Experiment 1 found little evidence that people differentiated between the epistemic weight given to independent and sequential evidence in a belief updating task. One concern about this null finding is that we did not check people's understanding of the respective evidence conditions, leaving open the possibility that individuals may not have interpreted the independent or sequential testimonies in the intended way. To address this concern, we conducted a follow-up study. This study replicated the five- and seven-informant independent and sequential testimony conditions used in Experiment 1, but added a more extensive probe of the participant's understanding of the independent/sequential manipulation. The full details of the experimental procedure and results for this follow-up study can be found at <https://osf.io/qng93/>. After completing a balls-and-urns task that was identical to that used in Experiment 1, participants in the follow-up study were asked the following forced-choice question: "You saw [5/7] friends make guesses about which jar was selected. Which of these statements about those guesses is most accurate?" The response options are shown in the following order:

- The [5/7] friends did not hear the other friends' guesses and did not see the balls given to the other friends [*correct response for independent condition*].
- The [5/7] friends could hear the other friends' guesses but did not see the balls given to the other friends [*correct response for sequential condition*].
- The [5/7] friends could hear the other friends' guesses and could also see the balls given to the other friends.
- The [5/7] friends saw a single ball and then agreed on the best guess.

The distribution of responses across these alternatives differed significantly between the independent and sequential conditions ( $\chi^2(3) = 33.13, p < .001$ ). In the independent condition, most participants (79%) responded correctly to the manipulation check question. Accurate responding to this question was less common in the sequential condition (46% of participants correct), but the correct option was still selected more frequently than any other option (see <https://osf.io/qng93/> for full details).

The pattern of likelihood ratings in the balls-and-urns task for participants in this follow-up study ( $N = 195$ ) was very similar to the empirical ratings shown in Fig. 3 (see <https://osf.io/qng93/> for a data summary). Participants in this follow-up study updated their likelihood ratings in response to new testimonial evidence ( $BF_{10} > 1000$  for five- and seven-informant conditions). However, this updating was not sensitive to the manipulation of evidence dependence in either the majority evidence ( $BF_{01} = 24.43, 34.16$  for five- and seven-informant conditions, respectively) or uptick stages ( $BF_{01} = 6.21, 5.46$  for five- and seven-informant conditions, respectively). Crucially, the null effect of evidence dependence was maintained when the analysis was restricted to participants who gave the correct response in the manipulation check ( $N = 122$ , majority evidence stage  $BF_{01} = 8.95, 16.11$  for five- and seven-informant conditions, respectively; uptick stage  $BF_{01} = 1.45, 4.26$  for five- and seven-informant conditions, respectively).

These results show that, although understanding of the independent/sequential manipulation was far from perfect, this manipulation did lead to different levels of belief in the dependence of the evidence presented in the urns task. Further, the null effect of dependence on belief updating was not due to a misunderstanding of this evidence manipulation.

### 2.3. Discussion

Experiment 1 and the follow-up study found that people revised their beliefs about the source urn in the expected directions as successive testimonies and/or own-balls were presented. The qualitative pattern of belief revision as new evidence was presented was generally consistent with the predictions of the Whalen et al. (2018) model. However, there was no difference in patterns of belief updating between independent and sequential testimony conditions, contrary to the Bayesian model predictions.

A key advantage of comparing empirical ratings with model predictions was that it sheds more light on why people failed to differentiate between independent and sequential testimony. Our model-fitting analyses suggested that many people underweighted the value of independent evidence, leading to likelihood ratings that were quite similar to those in the sequential conditions. This was true across the three-, five-, and seven-informant conditions.

Such conservatism in response to independent evidence has been well documented with the balls-and-urns task. Phillips and Edwards (1966) showed that participants often did not update to the extent predicted by Bayesian models after receiving independent ball draws. One reason that people may be conservative is because our everyday environments rarely contain true independence between successive observations (Navon, 1978; Winkler & Murphy, 1973). Thus, underestimating the level of independence may be adaptive (see Navon, 1978, for formal proofs supporting this claim).

It is unclear from this study, however, whether such conservative responding reflects a general propensity to doubt the assumption of data independence or whether it reflects limitations on the methods used to measure belief updating. Previous work with non-social versions of the balls-and-urns task has suggested that conservative belief updating may reflect the lack of incentive to use extreme ends of the response scale or the use of a linear rather than logarithmic scale for intuitive probability judgments (Phillips & Edwards, 1966). These issues were examined in Experiment 2.

### 3. Experiment 2

This experiment tested procedural explanations of the conservative updating found in the independent testimony condition of Experiment 1. We again compared belief updating following independent and sequential testimony but added financial incentives for accurate belief estimation and a logarithmic response scale for rating the likelihood of the source urn. These innovations were inspired by previous work using “non-social” versions of the balls-and-urns task.

Phillips and Edwards (1966) suggested that participants may update their beliefs in a Bayesian manner, but may be conservative in the way they express this in numerical ratings, showing reluctance to use the extremes of a response scale when warranted. Adding incentive payoffs for “accurate” updating (i.e., in line with Bayesian predictions) reduced conservative responding, although it did not entirely eliminate it. Hence, in the current Experiment 2, participants were told they could double their earnings if their final likelihood rating closely approximated Bayesian predictions.<sup>3</sup>

A second reason why participants’ likelihood ratings may not reflect underlying beliefs about the strength of the evidence relates to the type of response scale used. Phillips and Edwards (1966) tested different response scales and found that participants made judgments that were closer to Bayesian predictions when they used a log-odds scale, compared to those who used verbal odds, log probabilities, or probabilities on a linear scale.

In addition, we introduced practice trials before the experimental trials. The practice trials provided direct experience with the stochastic nature of the balls-and-urns task and showed how minority-colored balls could be drawn. This should help participants understand why an informant in a sequential cascade might provide a testimony that did not match the ball they had received.

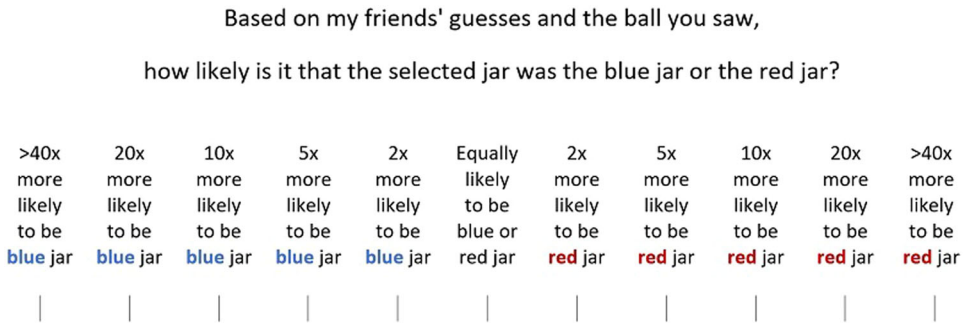


Fig. 4. Response scale used in Experiment 2. The example shown here is the final rating taken after all informant testimonies, and the participant’s own private ball were presented. The same scale, omitting reference to “the ball you saw,” was presented after each testimony. The increments of this 11-point scale were approximately consistent with a log scale.

### 3.1. Method

#### 3.1.1. Participants

Participants were 206 US residents recruited from MTurk. We excluded 12 participants who failed an attention check question, leaving a final sample size of 194 (mean age = 35.34,  $n = 62$  female,  $n = 132$  male).

#### 3.1.2. Design and procedure

The study used a 2 (dependency: independent, sequential)  $\times$  2 (number of informants: 5, 7) between-subjects design. All participants were incentivized, and all responded on the log-odds scale. We did not run a three-informant condition because, when converted to a log-odds scale, the Bayesian model predictions for the independent and sequential conditions only differed when there were more than three informants providing evidence— see <https://osf.io/qng93/>. The general procedure was similar to Experiment 1, but with three key differences.

First, after task instructions, participants completed two practice trials. On each trial, participants were shown a grayed-out jar (that could not be discerned as either the blue or red jar) and were prompted to draw a ball by clicking an on-screen button. A randomly sampled ball was then shown, according to the ball proportions known to the participant (i.e., 5/6 majority-color ball, 1/6 minority-color ball). This process was repeated for 12 balls. Then, participants provided a likelihood rating (scale described below), before being shown which jar had been selected. An alternative jar was selected for the second trial, which proceeded in the same manner. Note that these practice trials did not include any informant testimonies.

Second, a log-odds scale replaced the linear response scale used in Experiment 1. The scale used (Fig. 4) was similar to that used by Phillips and Edwards (1966). The scale had 11 points like the likelihood rating scale used in Experiment 1, and the increments between points were approximately consistent with a log scale (i.e., more sensitive at smaller values).

The third change was the addition of a monetary incentive for responding in line with Bayesian predictions. After the practice trials, participants were told about the opportunity to

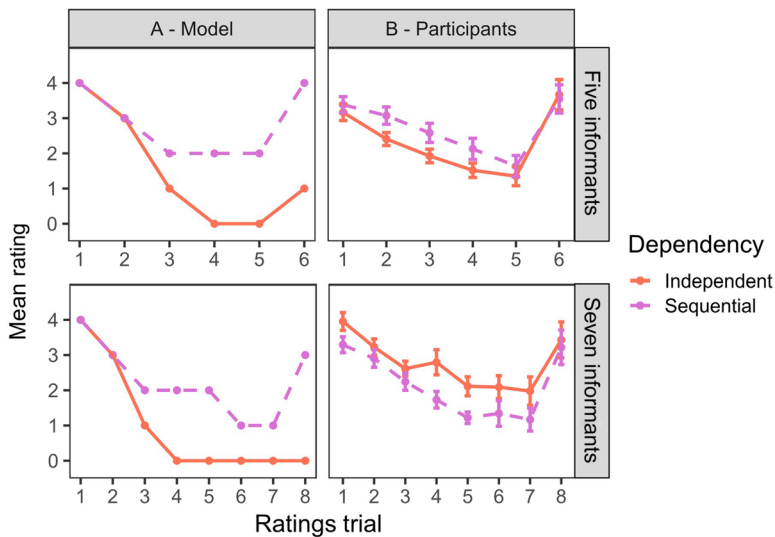


Fig. 5. Experiment 2 model predictions (A) and participants' mean ratings about the source urn (B) model predictions and mean ratings of the likelihood that the source urn matched the participant's private evidence after each testimony/ball draw.

win a bonus payment. The instructions stated that we would look at their judgments across the task, and “If at the end of the task, you have given the correct rating, or one rating either side, we will pay you an additional \$0.90” (see <https://osf.io/qng93/> for full instructions).

### 3.2. Results and discussion

Mean likelihood ratings and model predictions are shown in Fig. 5. To generate model predictions for this new response scale, we took the posterior probabilities from the Whalen et al. (2018) model, calculated the odds ratio, then selected the point on the 11-point response scale that was closest to that odds ratio. These calculations are available at <https://osf.io/qng93/>. The model predictions using this log-odds scale generate larger dependency effects compared to the original model predictions on the linear scale.

We ran repeated-measures Bayesian ANOVAs to assess participants' ratings during the testimony stage, and Bayesian  $t$ -tests to assess ratings following the final own-ball. Separate Bayesian ANOVAs were run for the five-informants and seven-informants conditions.

With five informants, there was inconclusive evidence against the main effect of dependency ( $BF_{01} = 1.03$ ) and strong evidence against an interaction effect between dependency and ratings trial ( $BF_{01} = 69.04$ ). Hence neither overall mean ratings nor the steepness of slopes of the likelihood ratings differed between independent and sequential conditions. The Bayesian  $t$ -test showed that the upticks following the final own-ball also did not differ between five-informant independent and sequential conditions ( $BF_{01} = 3.79$ ).

With seven informants, there was strong evidence supporting the main effect of dependency ( $BF_{10} = 82.55$ ) during the testimony stage. However, Fig. 5 shows that the effect was in the

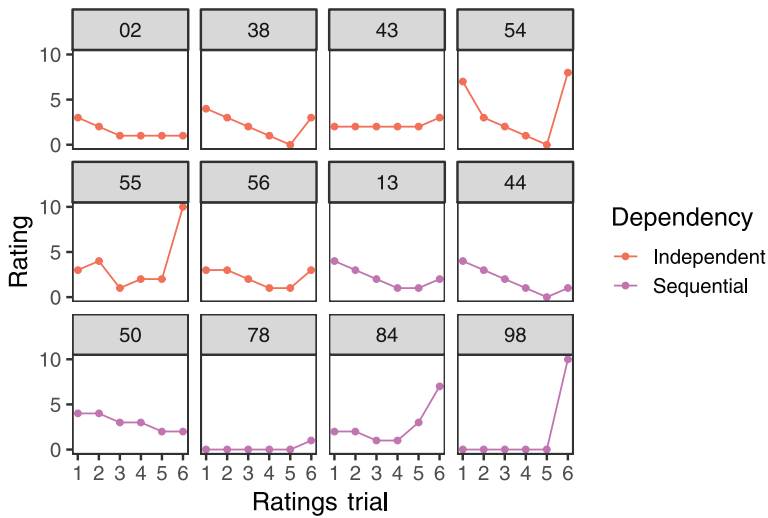


Fig. 6. Ratings from 12 individuals in Experiment 2 illustrate that response patterns often differed widely within each dependency condition.

*opposite* direction to model predictions—on average, independent ratings were *higher* than sequential ratings, when they should have been lower. We speculate that participants grew suspicious of seven identical but supposedly independent, testimonies. Seven informants each receiving the same majority-colored ball is relatively unlikely under a binomial distribution ( $p = .28$ ). Moreover, people often perceive long “streaks” of identical outcomes as evidence of a non-random data-generating process (Burns & Corpus, 2004). Thus, participants may have (over-)corrected for perceived dependence in the seven-informant independent conditions. As with five informants though, the final upticks following the own-ball were similar between independent and sequential conditions ( $BF_{01} = 2.98$ ).

As in Experiment 1, participants’ ratings were closer to model predictions in the sequential condition ( $R^2 = .83$ ) than in the independent condition ( $R^2 = .33$ ), aggregated across number-of-informant conditions. Again, Fig. 5 suggests that the lack of a difference in belief updating between dependency conditions was mostly due to conservative responding in the independent conditions.

In sum, we again found no clear evidence that participants updated beliefs differently in response to independent testimony compared to sequential social testimony, even when participants completed practice trials, were incentivized, and provided with a log-odds response scale. These data reinforce the view that people often find it difficult to see the difference in the epistemic value of independent and sequential social evidence.

As in Experiment 1, the model-based regression analyses suggested that belief updating in the independent evidence condition showed conservative updating relative to the predictions of the Bayesian model. It is important to remember, however, that up to this point, such conclusions have been based on judgments aggregated across participants. This aggregation obscures individual variability in belief updating patterns. Fig. 6 shows a sample of individual



likelihood ratings from Experiment 2, which were chosen from the five-informant conditions to illustrate the variability of responses. These profiles highlighted that, within dependency conditions, identical testimonies could lead to a variety of individual patterns of belief updating. For example, participant 38 updated with each successive independent testimony, while participant 43 seems uninfluenced by later testimonies. In contrast, participant 44 updated with every subsequent sequential testimony.

Given such within-group variability, Experiment 3 examined the possibility that subgroups of participants may have different prior beliefs about the relative value of independent and dependent evidence and this may influence whether individuals update their beliefs in a Bayesian manner.

#### 4. Experiment 3

The Bayesian model of belief updating assumes that (a) people recognize the difference in the epistemic value of independent and dependent social evidence and (b) use a process analogous to Bayes' rule to update their beliefs as independent or dependent evidence is encountered. This study examines the possibility that the lack of sensitivity to information dependency in the previous studies was because the first of these assumptions does not hold for a substantial proportion of our participants. That is, some participants do not believe that independent evidence is indeed more valuable than dependent evidence.

Relatively little previous work has directly probed participant beliefs about the relative value of independent and dependent evidence for judgments under uncertainty. However, there is some evidence that suggests heterogeneity in such beliefs. Pilditch et al. (2020), for example, presented participants with plane crash reports from two investigators, who both concluded that a plane was sabotaged. These corroborating reports were either completed sequentially, with the second investigator completing their report after accessing the first investigator's report, or independently, so that both investigators completed their reports without seeing the other report. When participants were asked which set of reports provided more support for the plane being sabotaged, just under 50% thought the independent case was stronger. Around 40% of participants believed that both forms of evidence were equally strong, while around 10% thought the dependent, sequential case was stronger.

In a similar vein, Yousif et al. (2019) presented learners with multiple statements supporting some claims (e.g., a new tax policy will benefit the Swedish economy). These statements came from secondary sources (e.g., newspapers) either reporting on a different, independent primary source each ("true consensus") or secondary sources that all relied on the same primary source ("false consensus"). When participants were asked which type of consensus would be more "believable," 50% of participants favored true consensus, with 16% favoring false consensus, and the remaining 33% seeing true and false consensus as equally believable.

It is possible, therefore, that many individuals completing our social balls-and-urns judgment task may have had a prior belief that sequential testimony should carry as much (or possibly more) weight as independent testimony. This could explain why the aggregated

likelihood judgments of those in the independent conditions of Experiment 1 were not distinguishable from those in the sequential condition.

To address this possibility, in Experiment 3 participants were given a more detailed explanation of the difference between independent and sequential social evidence before choosing which type of evidence they believed to be more convincing. They were also asked to justify this choice. We then manipulated information dependency within subjects; all participants completed both an independent and sequential version of the balls-and-urns task. This allowed us to examine individual sensitivity to dependency in belief updating and to relate such sensitivity back to individuals' stated beliefs about the value of independent information.

We expected that those who did not endorse independent over sequential information would show belief updating patterns similar to those found in the previous experiments. However, those who explicitly endorsed independent evidence as more valuable were expected to update their beliefs in a manner more consistent with Bayesian predictions—more updating following independent as compared with sequential testimony.

#### 4.1. Method

##### 4.1.1. Participants

Participants were 212 U.S. residents recruited from MTurk. We excluded 12 participants who failed an attention check question, leaving a final sample size of 200 (mean age = 37.04,  $n = 76$  women,  $n = 123$  men,  $n = 1$  non-binary). It was not possible to make an a priori estimate of the number of participants who would endorse the alternative options in our pre-test of preference for independent or sequential evidence. To allow for adequately powerful comparisons between these naturally occurring groups, we therefore recruited a larger sample than was used in the previous experiments.

##### 4.1.2. Design and procedure

Evidential dependency (independent vs. sequential) was manipulated within subjects, while the number of informants (three, five) was manipulated between subjects. We did not use the seven-informant condition because it was difficult to show seven example testimonies onscreen (see below description of how testimonies were presented). Moreover, Experiment 2 suggested that participants doubted that seven independent informants would give the same testimony.

Except as noted below, the general procedure was similar to that used in the previous experiments. Prior to experimental trials, participants completed practice trials like those in Experiment 2. The response scale was the 0–11 likelihood rating scale used in Experiment 1.

The two modifications to the procedure were as follows: (a) interrogation of participants' prior beliefs about the value of independent versus sequential testimony and (b) within-subjects manipulation of independent and sequential versions of the balls-and-urns tasks. Before completing the balls-and-urns task, participants were given an additional explanation of independent versus sequential testimony. As illustrated in Fig. 7A, these instructions contrasted the lack of access to other friends' guesses in the independent case with the

(A)

We are interested in what people think about this sort of social information  
(i.e., the information provided by my friends).




We will now show you 2 types of social information, and ask you to choose which type you prefer.

**1. Sequential information**  
The friends guess one after the other, in the same room.  
So each friend hears the previous friends' guesses, and uses that information to inform their own guess.

**2. Independent information**  
Each friend guesses in separate rooms, without knowing what the other friends guessed.

(B)

Here is an example of **sequential** information.

Mary	Jane	Sarah
		
"I looked at my ball and I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."	"I looked at my ball <i>and</i> I thought about what Mary said. I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."	"I looked at my ball <i>and</i> I thought about what Mary and Jane said. I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."

Here is an example of **independent** information.




Mary	Jane	Sarah
		
"I looked at my ball and I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."	"I looked at my ball and I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."	"I looked at my ball and I <b>guess the grey jar was selected.</b> I did not see any of the balls given to the others."

Fig. 7. Experiment 3 (A) instructions explaining independent versus sequential information and (B) examples of independent and sequential information provided to participants. After the general balls-and-urns task instructions, participants were given (A) extra comparative detail about independent and sequential testimony (in randomized order) and (B) examples of each type of testimony.

consideration of previous guesses in the sequential case. Next, participants saw each case illustrated with gray and white urns (different colors from those used in the main task), as shown in Fig. 7B. Each informant's testimony was revealed one at a time, accumulating on the screen, with all testimonies supporting the gray urn. The instructions did not indicate which type of evidence should be viewed as stronger or more convincing.

Both sequential and independent example testimonies remained on-screen while a question appeared at the bottom of the screen: "In your opinion, which set of social information provides stronger evidence for the gray jar being the selected jar?" The forced-choice response options were "Independent information," "Sequential information," or "Both

provide equal evidence.” Participants were asked to justify their response (“Why did you choose that answer?”), typing their answers into a text box.

All participants subsequently completed both an independent and a sequential version of the task, with task order randomized across participants. Different colored balls and urns (red/blue; purple/green) were used for each version. Assignment of ball/urn colors to versions was randomized for each participant. In between the two dependency versions of the task, instructions stated how the dependency would change (see <https://osf.io/qng93/> for verbatim instructions). The procedure for the respective independent and sequential testimony tasks followed that of the analogous conditions in Experiment 1.

## 4.2. Results and discussion

### 4.2.1. Preference for independent versus sequential evidence

In the forced-choice question, 38.5% of participants chose “independent information” as the more convincing form of evidence ( $n = 77$ ), 36.5% chose “sequential information” ( $n = 73$ ), and 25% expressed no preference (“both equal”;  $n = 50$ ). These percentages might be seen as implying that participants chose randomly. However, the open-ended justifications indicated that choices were based on deliberation about the options—with different choices associated with different types of justifications. Those who endorsed “independent information” most often justified this with reference to avoiding the biases created by information dependency (e.g., “Because with this method, you can be sure that they did not influence one another and that there is no bias”). In contrast, those who endorsed “sequential information” often claimed that informant consensus improves evidence quality (e.g., “With Sequential you get more information as you go, which should lead to a better guess”). Those who endorsed “both” most often justified this on the basis of the similarity between claims based on the two types of testimony (e.g., “Because everyone arrived at the same conclusion”). These participants appeared to be employing a simple counting rule (e.g., three guesses for red = three guesses for red). All justifications, and details of interrater reliability in the classification of judgments, are available at <https://osf.io/qng93/>.

### 4.2.2. Aggregate likelihood ratings

For comparison with previous studies, we began with an analysis of aggregate mean likelihood ratings (i.e., not segregating according to the preferred types of evidence). As in previous studies, participants revised their beliefs about the source urn downwards with additional testimonies and upwards after receiving their own-ball, in both the three- and five-informant conditions (all  $BF_{10} > 1000$ ). The within-subjects manipulation of independent and sequential testimony, however, still failed to produce evidence of sensitivity to dependency in these aggregated data. There was moderate evidence favoring the null hypotheses for comparisons between likelihood ratings in the independent and sequential conditions in overall value ( $BF_{01} = 5.99, 2.79$  for three- and five-informant conditions, respectively). Likewise, there was moderate evidence favoring the null hypothesis of no difference between the slopes of likelihood ratings in independent and sequential testimony conditions ( $BF_{01} = 7.26, 9.64$  for three- and five-informant conditions, respectively). The size of the final uptick also

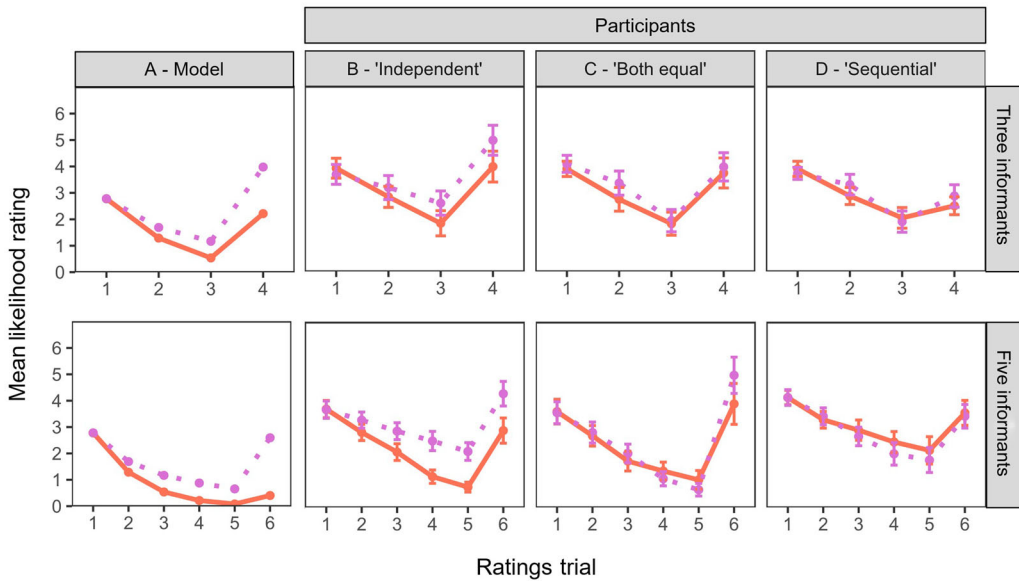


Fig. 8. Experiment 3 Model predictions and participants’ mean likelihood ratings. Model predictions and mean ratings of the likelihood that the source urn matched the participants’ private evidence after each testimony or ball draw. Mean likelihood ratings are shown for participants who judged independent information to be most convincing (B), judged independent and sequential information as equally convincing (C), or judged sequential information as most convincing (D). The two rows display the three- and five-informants conditions.

did not differ between independent and sequential conditions ( $BF_{01} = 5.02, 4.21$  for three- and five-informant conditions, respectively). This replication of the belief revision found in the earlier experiments is notable given the new procedural features in this study, including the within-subjects manipulation of dependency.

#### 4.2.3. Individual differences in preference for independent versus sequential evidence

The next set of analyses was similar to those above except that evidence preference was added as a between-subjects factor with three levels (prefer independent, prefer sequential, no preference). In three-informant conditions, there was moderate evidence against the main effect of evidence preference ( $BF_{01} = 45.65$ ) and against the interaction between evidence preference and dependency ( $BF_{01} = 21.17$ ). However, with five informants, there was very strong evidence for the main effect of evidence preference ( $BF_{10} > 1000$ ) and moderate evidence for an interaction between evidence preference and dependency ( $BF_{10} = 8.33$ ).

The right panels of Fig. 8 suggest that, in the five-informant condition, those who endorsed independent evidence were more likely to differentiate between independent and sequential evidence than other individuals. Follow-up analyses confirmed that these “independence endorsers” *did* differentiate between independent and sequential testimonies. These participants gave lower mean ratings overall ( $BF_{10} = 235$ ) and showed a steeper reduction in likelihood ratings ( $BF_{10} = 3.49$ ) following independent testimonies than sequential testimonies

(Fig. 8B).<sup>4</sup> Fig. 8B also shows that participants who endorsed independent testimony were less conservative when updating from independent testimony than other participants.

In contrast, those who preferred “sequential information” gave similar likelihood ratings in the independent and sequential conditions (all  $BF_{01} > 5.61$ ). Likewise, participants who chose the “both” option did not show sensitivity to evidential dependency—giving similar likelihood ratings in the independent and sequential conditions during the testimony stage (all  $BF_{01} > 6.52$ ).

The trends reported above were reinforced by a comparison of ratings given by each preference group to Bayesian model predictions. Linear regression analyses revealed that the responses of those who preferred “independent information” were closer to Bayesian model predictions ( $R^2 = .85$ ) than participants who selected “both” ( $R^2 = .60$ ) or “sequential information” ( $R^2 = .36$ ). To examine the robustness of these apparent group differences in model fit, we entered Whalen et al. (2018) model predictions and individuals’ evidence preference (independent, sequential, both) as predictors in a Bayesian linear regression. Because evidential dependency in the balls-and-urns task was manipulated within-subjects in this experiment, the dependent measure in the regression was the difference between participant likelihood ratings on each trial in the independent and sequential conditions. The Bayesian linear regression found that participant ratings were better explained by the regression model that included both model predictions and forced-choice preference (with three levels) as predictors, compared to the model without evidence preference ( $BF_{10} = 187$ ).

The finding that evidence preference did not affect likelihood ratings in the three-informant condition suggests that the number of informants, or the number of evidence trials more broadly, is important. Three informants appear to be too few for participants to distinguish independent from sequential testimonies, even when participants explicitly recognize the relative superiority of independent evidence.

Experiment 3 shows that sensitivity to evidential dependency in the balls-and-urns task depended on participants’ general beliefs about the relative value of independent and sequential information. Participants in the five-informant condition who endorsed independent information as the stronger form of evidence subsequently revised their beliefs in a manner that generally conformed to the predictions of the Bayesian model. By contrast, those who endorsed sequential evidence as stronger, or who expressed no preference, showed a pattern of belief revision similar to the aggregate data in our previous experiments—they did not differentiate between independent and sequential evidence when making intuitive likelihood judgments.

## 5. General discussion

The aim of the current work was to advance our understanding of the conditions under which people discriminate between the value of independent and dependent social evidence when making judgments under uncertainty. Across three experiments, we found that consistent with the qualitative predictions of a Bayesian model (Whalen et al., 2018), participants revised their beliefs about two possible states of the world as they received successive

testimonial and private evidence. However, contrary to the model predictions, most participants did not differentiate between the weight they attached to independent and sequential testimony when revising beliefs. This insensitivity persisted when there were as many as seven informants providing evidence (Experiment 1), when there were incentives for accurate performance and people could respond on a log scale (Experiment 2), and when there was a detailed explanation of the differences between the two types of evidence (Experiment 3).

A crucial exception was revealed in Experiment 3, where we examined people's beliefs about the relative value of independent and sequential evidence. A substantial minority (around 39%) expressed a general preference for independent information. Of these independence endorsers, those who heard from five informants showed more belief updating from independent testimony than from sequential testimony. In other words, participants whose prior beliefs aligned with the assumptions of the Bayesian model and were exposed to more than three informant testimonies produced likelihood ratings that were more consistent with Bayesian predictions.

### *5.1. Belief updating and preferences for independent or dependent evidence*

In the aggregated data from Experiments 1 and 2, we found that belief updating with independent evidence was conservative relative to Bayesian predictions. In these cases, those in the sequential evidence condition showed patterns of belief revision that were closer to the predictions of the Bayesian model. On reflection, the finding of conservative responding in the independent conditions may not seem too surprising, given that outside the laboratory, true independence between information sources may be rare (Corner, Harris, & Hahn, 2010; Navon, 1978; Winkler & Murphy, 1973). To return to our initial example, three seemingly independent tweets from independent news sources may ultimately share common data. Hence, the adaptive learner may adjust for assumed dependence, even in experimental contexts like ours where the instructions and experimental procedure were intended to convey independence (Connor Desai, Xie, & Hayes, 2022).

Crucially, however, Experiment 3 suggests that the apparent conservatism found in Experiments 1 and 2 (and similar results reported by Whalen et al., 2018) was most likely an artifact of averaging over participants with very different prior beliefs about the relative value of independent evidence. When participants in Experiment 3 expressed an a priori preference for independent over dependent social information, their subsequent responses in the belief-updating task were not overly conservative. This group showed updating in response to independent and dependent social testimony that generally conformed to the Whalen et al. (2018) Bayesian account.

Notably, this subgroup was in a minority. Nearly two-thirds of participants preferred dependent over independent evidence or expressed no preference. These participants updated their beliefs in the urn tasks after receiving social testimony but gave similar weight to dependent and independent testimonies.

This result reinforces previous findings that have found heterogeneity in peoples' beliefs about the value of independent over dependent social evidence (e.g., Pilditch et al., 2020;

Yousif et al., 2019). Our work, however, goes further than previous studies by showing how individuals' prior beliefs about the value of different types of evidence lead to different patterns of belief revision.

An important question that remains is the source of these different priors. One possibility is that individuals were prompted by our demonstration of the differences between independent and dependent evidence in the earlier stage of Experiment 3. That is, some participants may have learned about the value of independent evidence from this demonstration, rather than having this belief prior to the commencement of the experiment. According to this account, the "independence endorsers" were those who better understood the training demonstration and were better able to apply the relevant principles in the subsequent belief updating task. Of course, what still needs to be explained is why some participants learned more from these examples than others.<sup>5</sup>

Alternately, the justifications provided by participants in Experiment 3 for endorsing independent or sequential information suggest that, in some cases, individuals may have more enduring beliefs about the relative value of these types of evidence. Those who endorsed independent evidence often cited the advantage of receiving judgments from informants who were not influenced by others. In contrast, many of those who endorsed the sequential option believed that this provides "more information." That is, they saw sequential testimony as involving a combination of objective evidence in the form of the ball received by the informant and the informant's helpful judgment about the most likely source urn.

One way of examining the impact of such beliefs in future work would involve varying the explanations provided for why informants in the sequential chain report the same judgment as previous informants. One type of explanation could emphasize that such repetition is based on confidence in previous judgments (e.g., "Friends in the chain listen to the first friend because that friend has lots of experience with these tasks"). An alternative condition could explain the repetition of earlier judgments as being due to peer pressure (e.g., "Friends in the chain listen to the first friend because they want to fit in with their friend"). These alternate explanations should, respectively, increase or decrease reliance on sequential information in the belief updating task.

Differences in prior beliefs about the value of independent and dependent evidence may also be linked to established individual differences in the propensity to use social evidence when making judgments (Mesoudi, Chang, Murray, & Lu, 2015; Molleman, van den Berg, & Weissing, 2014). For example, in a stock investment game, Toelch, Bruce, Newson, Richerson, and Reader (2014) presented individuals with both a private hint (e.g., received a tip to invest in stock A) and a social hint (e.g., partner suggests investing in stock B). Although there were clear aggregate trends (e.g., people were more likely to copy the social hint when the private hint was known to be unreliable), some individuals relied on the social hint more often than others. Individual differences in the propensity to value independent evidence may also be related to other variables that have been shown to predict some forms of Bayesian reasoning, such as numerical ability (e.g., Brase & Hill, 2017) or cognitive reflection (e.g., Sirota et al., 2014).

Distinguishing between these various accounts of why people may have different beliefs about the relative value of independent and dependent evidence is a key goal for future



research. These accounts carry different implications for our ability to improve the way that people make judgments based on social evidence. If the individual differences we identified are primarily due to understanding of the training examples provided, then there are prospects for increasing sensitivity to evidence dependence through more extensive training. One method that may be worth trialing is to include participants as members of the information chain—giving them first-hand experience of how data are received and passed on in the independent and sequential scenarios. Related work has shown that asking participants to play the role of a generator of “fake news” increases subsequent sensitivity in distinguishing between factual and fictional media accounts (Roozenbeek & van der Linden, 2019).

### 5.2. *Implications for Bayesian models of belief revision*

Participant judgments in each experiment were compared against the normative standard provided by Whalen et al.’s (2018) Bayesian model of social learning. This model is useful in quantifying the uncertainty inherent in social testimony compared to directly observed private evidence as well as in quantifying how the redundancy in sequential testimony reduces its evidential weight compared to independent testimony. Although the “independence endorsers” in Experiment 3 integrated evidence in a manner consistent with this Bayesian model, a majority of participants did not.

Our findings suggest that the Whalen et al. (2018) Bayesian model only describes the beliefs and reasoning of a subgroup of participants (i.e., those who acknowledge that independent information should carry more weight). In contrast, when participants believe that sequential information is inherently more valuable, the Whalen et al. (2018) model does not apply. This is a good illustration of the general point that Bayesian models like Whalen et al. (2018) are better thought of as descriptive theories of how some individuals might reason rather than as normative or prescriptive models that apply to all (cf. Tauber, Navarro, Perfors, & Steyvers, 2017).

We also note that an alternative approach to modeling information cascades was suggested by Schöbel, Rieskamp, and Huber (2016), and that this model might provide a way of capturing the reasoning of people with different prior beliefs about the relative value of independent and sequential evidence. Like Whalen et al. (2018), their model follows Bayesian principles but incorporates an explicit “social importance” parameter that corresponds to the relative weight placed on social information compared with private information. To extend this model to the current studies, we would have to assume that the social influence parameter is set at a higher value for those individuals who endorse sequential over independent information.

A more general implication here is that future development of Bayesian models of judgment based on social evidence must consider individuals’ subjective causal models of how informant testimony was generated (cf. Hayes, Hawkins, Newell, Pasqualino, & Rehder, 2014; Szollosi & Newell, 2020). For example, the assumption of recursive Bayesian reasoning does not always seem to be appropriate: individuals who preferred sequential information did not seem to believe that later informants simply copied the judgments of earlier informants.

One caveat to our conclusions is that they may only apply to situations in which informants provide corroborating evidence. In these situations, the Bayesian model is clear—independent testimony provides more evidential weight than sequential testimony. However, there are cases where dependent testimony can become more valuable than independent testimony, such as when informants disagree (e.g., Pilditch et al., 2020). Future research could therefore examine how people's beliefs about independent versus sequential evidence, and their belief updating in the balls-and-urns task, compare with Bayesian model predictions in cases where dependence is expected to promote greater belief revision than independence.

### 5.3. *Generalizing to more complex social scenarios*

In our balls-and-urns task, we attempted to hold constant factors such as informant reliability and intentions by describing all informants as equally well-informed and helpful. We see this as a strength of the current design, in that it allows us to examine the effect of dependence between social information separately from considerations of informant expertise and motives. Outside the laboratory, however, these factors undoubtedly contribute to the way people interpret and use social information. When people believe that informants share the same information because they share a common motivation to persuade the listener (e.g., to buy a particular product), then this information is more likely to be discounted (Mercier & Miton, 2019). Madsen et al. (2019) have also shown that informants who repeat information provided by others may be perceived as less reliable or expert. Future work could therefore build on the current studies by systematically manipulating factors such as the motives and perceived expertise of the informants in independent and sequential evidence sequences.

## 6. Conclusions

Most Bayesian models of the judgment under uncertainty (e.g., Whalen et al., 2018) assume that we should give more weight to social testimony generated from independent sources than sources where the testimony of an informant depends on that of other informants. In three experiments using a social version of the balls-and-urns task, we found that although people do update their beliefs as they receive new testimony, they often showed similar patterns of updating based on independent and sequentially dependent evidence. A notable exception to these trends was found in a subgroup that expressed an explicit preference for independent over sequential testimony—and who subsequently showed more belief revision from independent evidence.

These results highlight the difficulty that people often experience in discriminating between the implications of independent social evidence and evidence subject to more subtle forms of social dependency. They also highlight the pressing need for future models of human judgment to incorporate individual differences in learners' assumptions about the implications of independent and dependent data. A failure to recognize such differences may lead to unnecessarily pessimistic conclusions about people's ability to adjust their judgments in response to dependent and independent evidence.

## Acknowledgments

We thank Chris Donkin for his advice on statistical analysis. This work was supported by the Australian Research Council Discovery Grant [DP190101224] awarded to Brett Hayes. Belinda Xie was supported by an Australian Government Research Training Program Scholarship and the UNSW Scientia PhD Scholarship Scheme.

Open access publishing facilitated by University of New South Wales, as part of the Wiley - University of New South Wales agreement via the Council of Australian University Librarians.

## Notes

- 1 Equation 2 represents the probability matching version of this model. In the maximizing model, informants instead deterministically support the hypothesis with the greatest posterior probability. Whalen et al. (2018) found that the probability matching model provided better fit to participant data and we therefore assume a probability matching model from hereon. Adopting the maximizing model would result in larger predicted differences between independent and sequential conditions, but the same qualitative patterns (see <https://osf.io/qng93/>).
- 2 We deemed these “medium” width priors to be more suitable than the other “wide” and “ultrawide” options available, because model predictions and Whalen et al.’s (2018) empirical findings suggested that any effect of dependency on likelihood ratings would be relatively small. We did however test the robustness of the results across priors. We repeated the analyses that examined a main effect of dependency, using a range of smaller  $r$  values corresponding to narrower priors. This generally weakened the evidence in favor of the null hypothesis but in no case did it result in evidence for the alternative hypothesis. Likewise, analyses that favored the alternative hypothesis with the default prior never favored the null hypothesis with narrower priors. Full details are available at <https://osf.io/qng93/>
- 3 Phillips and Edwards (1966) provided a formal account of why payoffs combined with a log-odds scale would increase Bayesian responding. However there is a less formal reason for expecting that adding payoffs for correct responding may lead to more extreme probability estimates. In the current design, participants were not given explicit feedback about what constituted a “correct” response. Given that only one urn could have been selected, they may nevertheless have intuited that the correct response had to be at one of the extreme ends of the response scale. We thank an anonymous reviewer for pointing this out.
- 4 Given the novelty of this finding, we felt a replication was required. In a follow-up experiment with  $N = 190$  online participants who had not participated in any of our previous experiments, participants again showed distinct preferences for independent and/or sequential testimony. In a subsequent balls-and-urns task, only those who endorsed the value of independent information showed different patterns of belief updating to inde-

pendent and sequential testimony, in line with Bayesian model predictions. Details are available at <https://osf.io/qng93/>.

- 5 A related argument is that, after expressing a preference for independent or sequential evidence, participants felt obligated to respond in a consistent manner in the belief updating task. While this is possible, we also found clear exceptions such as in the three-informant condition. In this case, some participants expressed a preference for independent evidence but did not show evidence of giving more weight to that evidence in the subsequent urns task.

## Open Research Badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://osf.io/qng93/>.

## References

- Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review*, 87(5), 847–862.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797–817.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026.
- Bloomfield, R., & Hales, J. (2009). An experimental investigation of the positive and negative effects of mutual observation. *The Accounting Review*, 84(2), 331–354.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298.
- Brase, G. L. (2021). Which cognitive individual differences predict good Bayesian reasoning? Concurrent comparisons of underlying abilities. *Memory & Cognition*, 49(2), 235–248.
- Brase, G. L., & Hill, W. T. (2017). Adding up to good Bayesian reasoning: Problem format manipulations and individual skill differences. *Journal of Experimental Psychology: General*, 146(4), 577–591.
- Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: Gambler's fallacy versus hot hand. *Psychonomic Bulletin & Review*, 11(1), 179–184.
- Connor Desai, S., Xie, B., & Hayes, B. K. (2022). Getting to the source of the illusion of consensus. *Cognition*, 223, 105023.
- Corner, A., Harris, A. J., & Hahn, U. (2010). Conservatism in belief revision and participant skepticism. In S. Ohlsson, & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, volume 32 (pp. 1625–1630). Austin, TX: Cognitive Science Society.
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
- Dere, M., & Boyd, R. (2018). Social information can potentiate understanding despite inhibiting cognitive effort. *Scientific Reports*, 8(1), 9980.
- Dietrich, F., & Spiekermann, K. (2013). Independent opinions? On the causal foundations of belief formation and jury theorems. *Mind*, 122(487), 655–685.
- Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., Spiegelhalter, D., & van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. *Journal of Risk Research*, 23(7-8), 994–1006.
- Enke, B., & Zimmermann, F. (2017). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1), 313–332.

- Hahn, U., von Sydow, M., & Merdes, C. (2019). How communication can make voters choose less well. *Topics in Cognitive Science*, 11(1), 194–206.
- Hayes, B. K., Hawkins, G. E., Newell, B. R., Pasqualino, M., & Rehder, B. (2014). The role of causal models in multiple judgments under uncertainty. *Cognition*, 133(3), 611–620.
- Hayes, B. K., Navarro, D. J., Stephens, R. G., Ransom, K., & Dilevski, N. (2019). The diversity effect in inductive reasoning depends on sampling assumptions. *Psychonomic Bulletin & Review*, 26(3), 1043–1050.
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025.
- Madsen, J. K., Hahn, U., & Pilditch, T. (2019). Reasoning about dissent: Expert disagreement and shared backgrounds. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2235–2242). Montreal, QB, Canada: Cognitive Science Society.
- Mercier, H., & Miton, H. (2019). Utilizing simple cues to informational dependency. *Evolution and Human Behavior*, 40(3), 301–314.
- Mercier, H., & Morin, O. (2019). Majority rules: How good are we at aggregating convergent opinions? *Evolutionary Human Sciences*, 1, e6.
- Mesoudi, A., Chang, L., Murray, K., & Lu, H. J. (2015). Higher frequency of social learning in China than in the West shows cultural variation in the dynamics of cultural evolution. *Proceedings of the Royal Society B: Biological Sciences*, 282(1798), 1–7.
- Molleman, L., van den Berg, P., & Weissing, F. J. (2014). Consistent individual differences in human social learning strategies. *Nature Communications*, 5(1), 3570.
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor. R package version 0.9.12-2.
- Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*, 279(1729), 653–662.
- Navarro, D. J., Dry, M. J., & Lee, M. D. (2012). Sampling assumptions in inductive generalization. *Cognitive Science*, 36(2), 187–223.
- Navon, D. (1978). The importance of being conservative: Some reflections on human Bayesian behaviour. *British Journal of Mathematical and Statistical Psychology*, 31(1), 33–48.
- Otsubo, K., Whalen, A., & Buchsbaum, D. (2017). Investigating sensitivity to shared information and personal experience in children's use of majority information. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 2840–2845). Austin, TX: Cognitive Science Society.
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163.
- Perreault, C., Moya, C., & Boyd, R. (2012). A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior*, 33(5), 449–459.
- Peterson, C. R., Schneider, R. J., & Miller, A. J. (1965). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology*, 69(5), 522–527.
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, 72(3), 346–354.
- Phillips, L. D., Hays, W. L., & Edwards, W. (1966). Conservatism in complex probabilistic inference. *IEEE Transactions on Human Factors in Electronics*, 7(1), 7–18.
- Pilditch, T. D., Hahn, U., Fenton, N., & Lagnado, D. (2020). Dependencies in evidential reports: The case for informational advantages. *Cognition*, 204, 104343.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163.
- Ransom, K., Hendrickson, A. T., Perfors, A., & Navarro, D. J. (2018). Representational and sampling assumptions drive individual differences in single category generalisation. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 930–935). Austin, TX: Cognitive Science Society.
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T., & Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science*, 328(5975), 208–213.

- Roozenbeek, J., & van der Linden, S. (2019). The fake news game: actively inoculating against the risk of misinformation. *Journal of Risk Research*, 22(5), 570–580.
- Schöbel, M., Rieskamp, J., & Huber, R. (2016). Social influences in sequential decision making. *PLoS ONE*, 11(1), e0146536.
- Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review*, 21(1), 198–204.
- Soll, J. B. (1999). Intuitive theories of information: Beliefs about the value of redundancy. *Cognitive Psychology*, 38(2), 317–346.
- Szollosi, A., & Newell, B. R. (2020). People as intuitive scientists: reconsidering statistical explanations of decision making. *Trends in Cognitive Sciences*, 24(12), 1008–1018.
- Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, 124(4), 410–441.
- Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior*, 52(2), 145–161.
- Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2014). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B: Biological Sciences*, 281(1776), 20132864. Publisher: Royal Society.
- Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes*, 103(1), 37–52.
- Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, 160, 110–126.
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57.
- Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive Science*, 42(1), 168–187.
- Winkler, R. L., & Murphy, A. H. (1973). Experiments in the laboratory and the real world. *Organizational Behavior and Human Performance*, 10(2), 252–270.
- Xie, B., Navarro, D. J., & Hayes, B. K. (2020). Adding types, but not tokens, affects property induction. *Cognitive Science*, 44(e12895), 1–38.
- Yousif, S. R., Aboody, R., & Keil, F. C. (2019). The illusion of consensus: A failure to distinguish between true and false consensus. *Psychological Science*, 30(8), 1195–1204.