Belief revision in a micro-social network: Modeling sensitivity to statistical dependencies in social learning

Anonymous CogSci submission

Abstract

Both in professional domains and everyday life, people frequently rely on their "social network neighbors" to form and update their beliefs. Here, it is important to understand how people deal with statistical dependencies underlying correlated beliefs in their social environment. Using an interface allowing us to elicit full probabilistic beliefs from people, we investigated people's ability to distinguish between the evidential value of social information across three conditions: integrating independent beliefs, dependent beliefs formed on the basis of the shared evidence, and dependent beliefs that result from sequential communication between sources. Comparing participants' judgments to a normative Bayesian model, we found that they distinguished dependent from independent sources but treated social sources as much weaker sources of evidence than direct experience. The value of eliciting and visualising beliefs as full probability distributions and potential implications for modeling belief revision in social networks (e.g., using agent-based models of echo chambers) are discussed.

Keywords: social networks; probabilistic beliefs; sequential belief updating; information cascades; Bayesian modeling

Introduction

We live and learn in a "society of minds" (Minsky, 1988). This means that we form beliefs, not just on the basis of our own observations (and prior expectations), but also based on the beliefs communicated by our neighbours in our vast distributed social network. When sitting on an interview panel, we discuss job applicants amongst ourselves even though we all have access to the same application materials, and were present for the same interviews, because this process is normally beneficial in updating our initial beliefs about applicants. Similarly, imagine you have read about two political campaign strategies, each proposed by a different candidate. If initially you find both strategies equally compelling, resulting in uncertainty about which of the two candidates you would like to support, you might well seek out new information about the candidates by talking to friends. If your friends based their beliefs partially on reading the same article, how do you weigh their opinions of the two candidates given that they are based on the same shared information?

The above examples illustrate one possible statistical dependency between opinions in a social network (i.e. shared information). If we are to understand learning in a social world, we must understand how people deal with such statistical dependencies while integrating their direct observations from the environment with the communicated beliefs of their social network neighbours. Investigating how information spreads through social networks and how statistical dependencies affect the formation of people's beliefs is thus a key issue for cognitive science with implications for e.g., the study of misinformation and echo chambers (Bikhchandani, Hirshleifer, & Welch, 1992; Watts, 2002; Whalen, Griffiths, & Buchsbaum, 2018; Madsen, Bailey, & Pilditch, 2018), the dynamics of micro-targeting (Madsen & Pilditch, 2018), or advocacy organisations' attempts to shape public debate (Bail, 2016).

Here, we investigated how people integrate information based on statistical (in)dependencies underlying the beliefs of three social network neighbours. We first introduce a Bayesian model of belief revision to account for the normative case. Building on previous work on sensitivity to shared information in social learning (Whalen et al., 2018), we compare three different conditions, the first serving as baseline in a sense that statistical independence between the beliefs of network neighbours A-C is induced by the cover story (see Fig. 1). In the second condition, independence is violated as participants are told that the three network neighbours form their beliefs on the basis of shared information (i.e. overlapping data). In a third condition, social network neighbours are described as having formed their beliefs sequentially, such that the belief of the neighbour that formed their belief last (neighbor C) contains all information gathered by the others. Dependencies between beliefs in condition three differ from dependencies in condition two in the sense that neighbor C is the only relevant source for updating one's belief based on the beliefs of the group. Shared information does not necessarily imply that other neighbors are redundant, but in contrast to the independent case, the aggregate of their beliefs is less informative (see Normative Framework). We report on a behavioural experiment that investigated how subjects update prior beliefs under these conditions.

Information Cascades and Probabilistic Beliefs Previous research has suggested that information cascades —spreading of beliefs through networks—produce maladaptive collective outcomes even as agents incorporate information from network neighbours in individually rational ways (Bikhchandani et al., 1992). This inherent susceptibility of social networks towards information cascades has also been supported by simulation-based research demonstrating that information cascades occur in networks of stochastic reinforcement learners equipped with

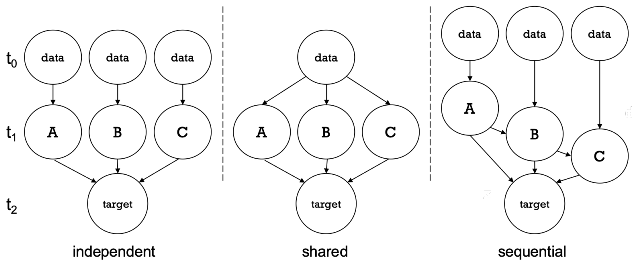

Figure 1: Illustration of network conditions. t_0: neighbors form beliefs given data. t_1: neighbors update beliefs based on interaction (sequential case only). t_2: target updates belief.

an individually rational cognitive architecture (Pilditch, 2017). Similar results have been obtained from empirical analyses of social media data, which demonstrate that users bundle in communities dominated by like-minded others, resulting in proliferation of unsubstantiated beliefs or conspiratorial thinking (Del Vicario et al., 2016).

In such previous models of information cascades, it is sometimes assumed that people settle on particular beliefs by maximising over subjective probabilities $p(b = h|d) = 1$ for $h = \text{argmax}_h\ p(d|h)p(h)$, leading to degradation of the information transmitted (Bikhchandani et al., 1992; Pilditch, 2017). Expanding this body of literature, recent empirical results have shown that people's social learning strategies are adaptive - accounting for statistical dependencies underlying the beliefs in their social networks, which might reduce the occurrence of information cascades (Whalen et al., 2018). In addition to maximizing, Whalen et al. (2018) tested the assumption of "probability matching", which assumes that people settle on beliefs stochastically, drawing particular conclusions proportional to the posterior probability of a belief $p(b = h|d) = p(d|h)p(h)$. In the present work, we make neither assumption (matching and maximizing), empirically exploring a setting in which agents communicate their full probabilistic beliefs. This allows us to explore the influence of communicated certainty in a belief. Certainty—defined as the inverse variance of the probability distribution encoding a belief—and its related probabilistic quantity confidence—the probability that a particular choice is correct— are at the core notion of an rational agent (Pouget, Drugowitsch, & Kepecs, 2016; Fleming & Daw, 2017), thus playing a crucial role during the integration of social information to update prior beliefs (see e.g., De Martino, Bobadilla-Suarez, Nouguchi, Sharot, & Love, 2017).

Normative Framework

We explore a general sequential belief updating process in which people first gather evidence by themselves (i.e. asocial information) to form an initial belief about two fictitious competing candidates. Collecting evidence means testing the candidates on two arbitrary performance tests of the form 0 = loss; 1 = win. We therefore model people's beliefs as beta probability distributions whose parameters are updated as more data becomes available. Following Bayes' rule, the initial posterior probability of a belief or hypothesis $p(h)$, given asocial information, d, thus corresponds to the normalised product of the likelihood $p(d|h)$ and the prior $p(h)$:

$$p(h|d) \propto p(d|h)p(h) \tag{1}$$

In our computational model, we use an uninformative beta-prior Beta(1, 1) which is conjugate to a binomial likelihood $\binom{n}{k}p^k(1-p)k^{n-k}$ with $n = 2$, allowing us to model belief updating straightforwardly using the analytical posterior $\text{Beta}(1+k, 1+n-k)$. For example, from an uninformative prior belief $X \sim \text{Beta}(1,1)$, observing data $D = \{0,1\}$ where k = 1, a person's posterior belief corresponds to $X \sim \text{Beta}(2,2)$. This reflects the nature of subjects' beliefs about the two candidates, which include a preference for candidates (mean of the beta distribution) and a measurement of certainty (inverse variance of the beta distribution). Following initial updating based on asocial information, we derive clear qualitative (directional) and quantitative predictions for how people should update their initial posterior beliefs upon observing the beliefs of other network neighbors between conditions. As in Equation 1, we use Bayes' theorem to model how people should integrate the beliefs (i.e. social information) from their network neighbors s:

$$p(h|s_1, \ldots s_n) \propto p(s_1, \ldots s_n|h)p(h) \tag{2}$$

Assuming that the beliefs of neighbors are independent, the final posterior distribution of a person's belief after incorporating the beliefs of social network neighbors is equal to aggregating the parameters of the neighbors and the target's initial posterior parameters. If neighbors are dependent in the sense that they updated their beliefs sequentially, our model predicts that the aggregated parameters of the posterior distribution should be based on the target's initial posterior parameters and the parameters of the neighbour that formed their belief last (i.e. neighbor C in Fig. 1). Finally, if sources are dependent in the sense that their beliefs are based on shared information, our model provides upper- and a lower bounds for the revised posterior. The upper bound is equal to the independent case, and it assumes that none of the neighbors' beliefs were influenced by the shared data (i.e. $D = \{\}$). Since we do not vary the parameters of neighbors between conditions in our experiment, the only source of variation in updating can be attributed to manipulating the dependencies between neighbors. Thus, the model lower bound is equal to subtracting the lowest α and β parameters from the aggregate parameters. For the present experiment, we do not provide a clear specification on how much neighbors were influenced by the shared data. Thus, assuming all possible combinations of overlap equi-probable, we model the normative impact of shared information on belief updating as having a magnitude intermediate between strictly independent information (higher magnitude) and sequentially updated beliefs (lower magnitude).

Based on this framework, we derived the following qualitative (directional) predictions: The difference between subjects initial- and revised posterior probabilistic beliefs will be smaller when the beliefs of social network neighbors are dependent as compared to the independent condition. Moreover, we predict that the dependent case of sequentially updated beliefs will result in a smaller update of prior beliefs as compared to shared information.

Experiment

Participants Participants (N = 79, range: 21 - 69 years, mean = 39.89, SD = 12.97, 35 female) were recruited and tested through Amazon's Mechanical Turk[1]. Participants were native English speakers based in the United States. They were paid \$1.75 for their time (mean = 17.49 min, SD = 6.91 min).

Task Description and Measures Participants imagined being a political consultant travelling around the US to help local branches of their political party decide between two fictitious competing candidates most suitable for public office. To do so, they imagined that they were travelling to three different cities, with two different candidates competing in each city. Prior to the main task, participants completed a short training phase and comprehension quiz to ensure that they understood how to provide their beliefs using the interface shown in Fig. 2. Specifically, participants used two response sliders to provide their full probabilistic beliefs (i.e. beta densities), one controlling the mean of the density (belief slider) and one controlling the log inverse variance (certainty slider; see Fig. 2). The response sliders ranged from 1-99; where a belief of 1 means full support for the left candidate, 50 is neutral, and 99 full support for the right candidate. A certainty of 1 is the lowest possible certainty and a certainty of 99 is the highest possible certainty. The resulting density was dynamically displayed (i.e. real time updating) to participants as they selected their response. Visualisations were restricted to concave function shapes (i.e. $\alpha \geq 1$ or $\beta \geq 1$).[2]

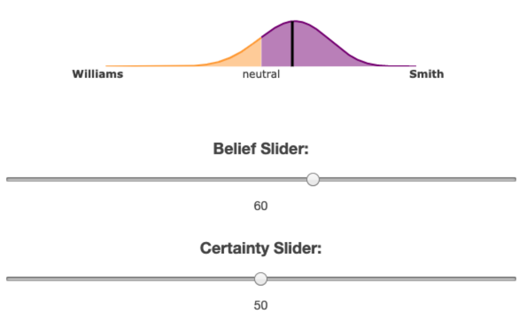

Figure 2: Interface for rating belief and certainty.

Within each city, the order of steps was: (1) obtaining prior belief based on observing asocial evidence $\rightarrow$ (2) learning about the beliefs of social network neighbors under consideration of statistical (in)dependence $\rightarrow$ (3) providing final posterior belief. We set the scene using an uninformative prior belief $X \sim \mathrm{Beta}(1,1)$ telling subjects that the two competing candidates were tested on two initial tests (each winning one of them) prior to the arrival of participants. Participants then observed the performance of the two competing candidates across two additional independent selection tests assessing different qualifications not covered in the initial tests. Following observation of asocial information, participants rated their prior belief and certainty in the relative suitability of the two competing political candidates for public office. The procedure was identical for each condition. After the initial assessment phase, subjects were shown the belief and certainty ratings of three social network neighbors (i.e. social information; see Fig. 3). The network neighbors were described as three locals that were likely voters from a subject's political party who learned about the candidates during debates in their local town-hall. Each city included a different cover story about the relationship between the three locals matching either statistical independence, shared information, or sequential belief updating. After learning about the beliefs of locals and their relationship, subjects provided their final belief.

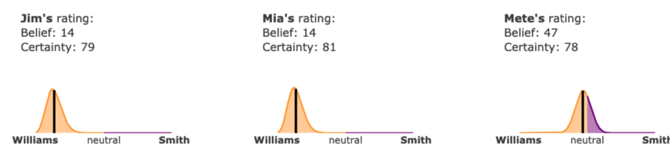

Figure 3: Example beliefs of locals.

Design and Procedure We employed a within-subjects design with three levels (variations of network setups implied by differing cover stories). The three levels of our independent variable were: 1) independent information; 2) shared information; and 3) sequential belief updating. The two data points (i.e. test outcomes) used to parameterize subjects' initial (neutral) prior beliefs, the resulting normative prior, and the parameters of the three locals are shown in Table 1. Parameter settings were constant across conditions, with the only source of variation being our independent variable. Resulting model predictions and sufficient statistics are summarised in Fig. 4 (columns 1-2). The order of conditions and the position of the candidate supported by locals (left/right) was randomized between participants. After completion of the main task, participants provided basic demographics (e.g., gender).

Table 1: Fixed parameters used across conditions.

Data	Normative Prior	Parameters of Locals
$\{1,0\}$	B(2,2)	B(46,9.5), B(49,10), B(50,45)

Analysis Our analysis has two parts. First, we compared the aggregate parameters of subjects' posterior judgments between

[1] OSF preregistration *here*.

[2] Full OSF folder including demo video of the experiment *here*.

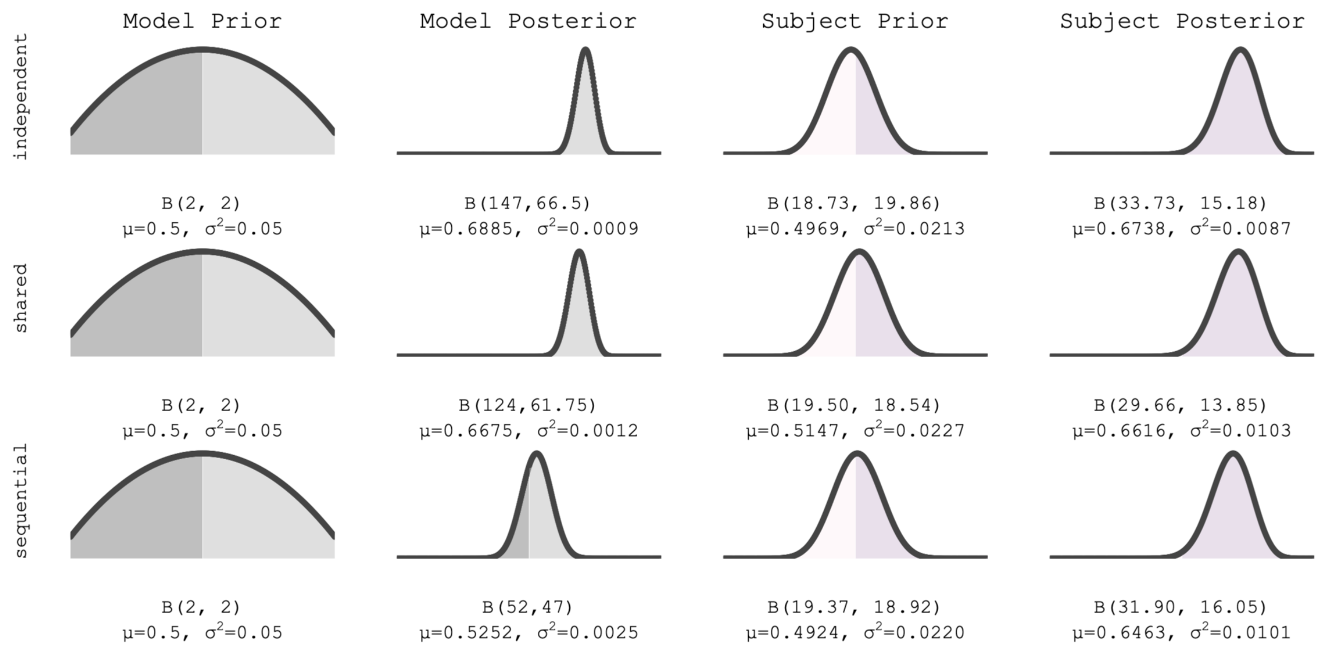

Figure 4: Summary of model predictions (columns 1-2) and behavioural results (columns 3-4) for each condition (rows 1-3). $B(\alpha, \beta)$ refer to the aggregate parameters of model predictions / subject responses for each condition and measure (i.e. prior and posterior). In our analysis, we used the aggregate μ and σ^2 parameters (plotted below $B(\alpha, \beta)$) to make the interpretation of our predictions and results more intuitive.

conditions to evaluate qualitative (i.e. directional) alignment with our model predictions. Thus, we first contrasted subjects' posterior means (model 1) and variances (model 2) between conditions using two linear mixed-effects models with condition (i.e. social network set-up) as fixed effect and subject as random effects. However, evaluating these separately may miss dependencies that exist in how participants updated these components of their beliefs. To address this, we computed the Jensen-Shannon Divergence (D_{JS}) between priors and posteriors for each subject to determine whether the magnitude of updating prior beliefs differed by condition. D_{JS} allows measurement of changes both in mean and variance between distributions through a single symmetric distance measure given by:

$$D_{JS}(P||Q) = \alpha D_{KL}(P||Q) + (1-\alpha) D_{KL}(Q||P) \quad (3)$$

where D_{KL} is the Kullback-Leibler Divergence, a standard asymmetric measure in information theory for measuring how much a probability density P has moved compared to a reference distribution Q. By definition, $D_{KL} \geq 0$, being equal to 0 if and only if P and Q are identical. A limitation of D_{KL} is its nonsymmetry, which is resolved by D_{JS} if $\alpha = 0.5$. Having computed each subject's D_{JS}, we fitted two additional mixed-effects models with condition as fixed effect and subject as random effect to compare mean differences in D_{JS} (model 3) and the log-transform of D_{JS} (model 4). The reason for using $\log D_{JS}$ in model 4 is that the distribution of $\log D_{JS}$ residuals was closer to a normal distribution than the distribution of D_{JS} residuals (which showed a skew to the right). All models were compared to a reduced model including only an intercept as predictor variable and subject as random effect. Models were implemented in R using the function `lmer()` from the package `lme4` (Bates, Mächler, Bolker, & Walker, 2015).

Following tests of our qualitative (directional) comparison between conditions, we compared subject and model performances across conditions to check in how far subjects aligned quantitatively with our normative framework. Therefore, we first computed D_{JS} between subjects' prior beliefs (Fig. 4, column 3) and the normative prior (Fig. 4, column 1) across conditions. Due to the skewed distribution of D_{JS} for this comparison, we conducted a Wilcoxon signed rank test (nonparametric t-test; alternative hypothesis > 0) to check if subjects integrated the asocial information as predicted by our model. We also compared the difference between subjects' prior mean and model prior mean and subjects' prior variance and model prior variance (using two-sided Wilcoxon signed rank tests, again because the dependent variables did not follow a normal distribution). The three comparisons were repeated to contrast subjects' posteriors with model posteriors.

Results

Sanity checks Five subjects were removed from our analysis because they did not change the positions of sliders between their prior and posterior judgments (i.e. $\log D_{JS}$ = -Inf), resulting in a final sample size of 74. Levene's test revealed that the homogeneity of variance assumption was maintained for all four dependent measures (all $ps > 0.05$) used between models 1-4. Inspection of residual plots confirmed that the residual posterior means, posterior variances and $\log D_{JS}$ residuals were normally distributed. For D_{JS}, residuals showed skew to the right. Correlations between the three levels of our fixed effect (i.e. social network set-up) were moderate, ranging from ± 0.470 to ± 0.551. Comparing each model to its reduced version revealed that inclusion of social network set-up only contributed significantly to the proportion of explained variance in $\log D_{JS}$ (see Table 2).[3] The qualitative comparison in the remainder of this paper will thus focus on interpreting the results of model 4 (for completeness, regression coefficients for models 1-3 are reported in the next section and in Fig. 5).

Table 2: Overview of model fits for each variable (DV).

Model	DV	BIC_{diff}	R^2_m	χ^2	p-value
1	μ	6.14	0.009	4.69	0.096
2	σ^2	9.10	0.004	1.65	0.439
3	D_{JS}	6.40	0.012	4.52	0.105
4	$\mathbf{\log D_{JS}}$	**-1.43**	**0.032**	**12.27**	**0.002**

Qualitative comparison Fig. 4 summarises model predictions and aggregate parameters across subjects for our three experimental conditions. The directional shift from subjects' prior distributions to their posteriors and the fact that subjects' posterior distributions were more compressed than their priors suggested that social information resulted in increased belief and confidence ratings. As expected by our normative model, the results of model 4 showed that subjects changed their prior beliefs significantly less in the two dependent conditions as compared to the independent case (Fig. 5D; b = -0.639, $t(148)$ = -3.07, p = 0.003 for sequentially updated beliefs and b = -0.641, $t(148)$ = -3.08, p = 0.003 for beliefs based on shared information). This means that the magnitude to which people updated their beliefs (i.e. changed both the belief and certainty sliders) was in line with the predicted (directional) magnitude of our normative model. In other words, people were sensitive to differences in the statistical power of the information between the independent condition (larger statistical power implying stronger updating of prior beliefs) and the two dependent conditions (smaller statistical power implying smaller updating). For the comparison between the two dependent conditions, no significant effects emerged (all $ps > 0.05$). The results of models 1 (μ) and 3 (D_{JS}) showed that the comparison between sequentially updated beliefs and independent beliefs was significant (b = -0.028, $t(148)$ = -2.16, p = 0.032 for model 1; Fig. 5A and b = -1.703, $t(148)$ = -2.02, p = 0.045 for model 3; Fig. 5C). For model 2 (σ^2), no significant differences emerged (all $ps > 0.05$; Fig. 5B).

[3] $BIC_{diff} = BIC_{full} - BIC_{intercept\text{-}only}$; R^2_m = proportion of variance explained by the fixed effect (i.e. social network set-up).

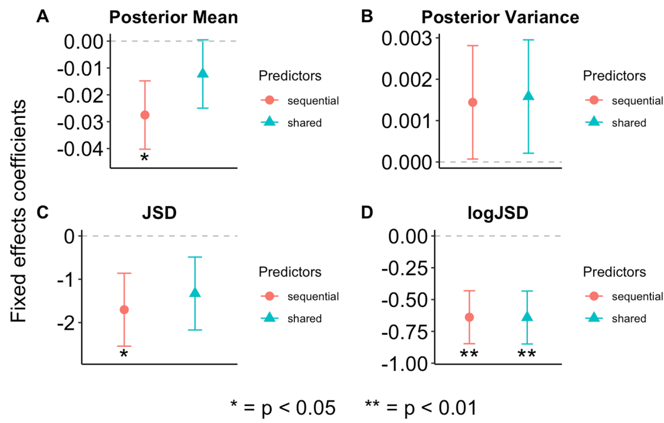

Figure 5: Coefficients for the two dependent conditions compared to the independent case (intercept in each model). Error bars correspond to standard errors of the mean.

Quantitative comparison Quantitative comparisons revealed that subjects' prior distributions differed significantly from model priors in terms of D_{JS} (V = 25425, $p < 0.001$). Inspection of Fig. 4 suggests that this difference might be driven by dissimilar variances, as subjects' priors were less diffused than model priors. The results of further comparisons confirmed this observation, showing that the mean of subjects' prior variance was significantly smaller than model prior variance (V = 1607, $p < 0.001$), despite finding no significant difference between prior means (V = 13150, p = 0.652). Means and standard deviations are shown in Table 3.

Table 3: Means and SDs of the dependent variables (DV) used for quantitative comparison.

Measure	DV	Mean	SD
prior	D_{JS}	2.547	3.782
prior	$\mu_{subj} - \mu_{model}$	0.001	0.085
prior	$\sigma^2{}_{subj} - \sigma^2{}_{model}$	-0.027	0.023
posterior	D_{JS}	6.270	8.491
posterior	$\mu_{subj} - \mu_{model}$	0.033	0.132
posterior	$\sigma^2{}_{subj} - \sigma^2{}_{model}$	0.008	0.012

Posterior contrasts revealed that model and subject distributions were significantly different from each other as measured by D_{JS} (V = 25425, $p < 0.001$). For posteriors, this difference was driven by a mismatch both in terms of posterior means (V = 17196, $p < 0.001$) and variances (V = 25183, $p < 0.001$). These findings demonstrate that, overall, subjects changed their posterior means more than expected by the normative prediction (mainly due to a strong mismatch between posterior means in the sequential belief updating condition, see Fig. 4, row 3). Moreover, subjects down-weighted social information provided by locals, which can be inferred from the fact that their average posterior variance was significantly larger than model posterior variance ($\sigma^2{}_{subj} = 0.01$; $\sigma^2{}_{model} = 0.002$).

Discussion and Further work

We modeled a sequential belief updating process including a target agent (i.e. the participant) and three social network neighbors. The cover story describing the relationship between network neighbors was varied in three within-subject conditions to investigate the effects of three statistical (in)dependencies summarised in Fig. 1. Extending the findings of Whalen et al. (2018), our behavioural results confirmed our prediction that people update their beliefs significantly less when the provided social information was coming from dependent sources (as compared to the independent case). Thus, our result shows that people are not simply combining their own beliefs with the communicated beliefs of their network neighbors. Rather, they are additionally sensitive to the origin of those beliefs and to what extent they are redundant. This contributes another empirical piece to the puzzle of how to mitigate the spread of false consensus effects and information cascades, which have been suggested to occur in networks of agents forming their beliefs in individually rational ways (Bikhchandani et al., 1992; Pilditch, 2017).

We could not confirm that people differentiated between the evidential signals of shared information and sequentially updated beliefs while revising prior beliefs (despite matching the normative pattern; see Fig. 4). This might be attributed to our task: social network neighbors were operationalized as locals being likely voters from a subject's political party that formed their beliefs based on attending debates in their local town-hall. A more abstract task, such as learning coordinating with others to estimate the proportion of blue vs. red marbles in an urn (a common paradigm used to study information cascades and sequential belief updating; see e.g., Anderson & Holt, 1997) or a formal explanation of how locals formed their beliefs (e.g., based on the same vs. different selection tests than the subject) might have resulted in a measurable difference. Despite being unable to differentiate between the two dependent cases, our task requiring people to update their beliefs in a noisy political context might be more ecologically valid than urn-based tasks, providing a valuable contribution to the field of social learning in the context of political belief formation (see e.g., Bond et al., 2012).

Contrary to our normative account, quantitative comparison between model predictions and subjects revealed that subjects over-weighted the influence of asocial information while they under-weighting the influence of social information. This finding is in line with previous theoretical (Schöbel, Rieskamp, & Huber, 2016) and empirical (Nöth & Weber, 2003) work demonstrating that people are more influenced by their own private information as compared to social information, though in our case this might be an artifact of simulated social information (i.e. information coming from hypothetical social network members rather than actual ones). In future work we plan to replicate the experiment in a setting where actual subjects come in pairs/triples to do the same task, enabling them to update beliefs dynamically. To generate a better understanding of whether people appropriately down-weight social evidence, further empirical work could incorporate agent-based simulations contrasting normative accounts with competing models of social influence (e.g., Schöbel et al., 2016). This approach could also make it possible to measure the empirical degree of information degradation due to correlated sources in realistic information networks. At the present stage, our finding suggests that people are able to understand and report probabilistic beliefs, which might be useful for calibrating belief parameters in related agent-based models (ABMs) of echo chambers (Madsen et al., 2018) and scientific belief formation (Lewandowsky, Pilditch, Madsen, Oreskes, & Risbey, 2019). Moreover, improving the cognitive sophistication of agents through a probabilistic representation of beliefs might also benefit related simulation work on coupled human-environmental systems (Bailey et al., 2019).

In summary, our results suggest that while a Bayesian framework provides a good qualitative account of how people update their beliefs based on social information coming from sources with different levels of independence, it cannot, in the current form, account for the relative weights that people assign to private and social information while updating their beliefs.

References

Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American economic review*, 847–862.

Bail, C. A. (2016). Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. *Proceedings of the National Academy of Sciences*, *113*(42), 11823–11828.

Bailey, R. M., Carrella, E., Axtell, R., Burgess, M. G., Cabral, R. B., Drexler, M., … Saul, S. (2019). A computational approach to managing coupled human–environmental systems: the poseidon model of ocean fisheries. *Sustainability Science*, *14*(2), 259–275.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

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.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., 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.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., … Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgments according to its reliability. *Journal of Neuroscience*, *37*(25), 6066–6074.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general bayesian framework for metacognitive computation. *Psychological review*, *124*(1), 91.

Lewandowsky, S., Pilditch, T. D., Madsen, J. K., Oreskes, N., & Risbey, J. S. (2019). Influence and seepage: An evidence-resistant minority can affect public opinion and scientific belief formation. *Cognition*, *188*, 124–139.

Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. *Scientific reports*, *8*(1), 12391.

Madsen, J. K., & Pilditch, T. D. (2018). A method for evaluating cognitively informed micro-targeted campaign strategies: An agent-based model proof of principle. *PloS one*, *13*(4), e0193909.

Minsky, M. (1988). *Society of mind*. Simon and Schuster.

Nöth, M., & Weber, M. (2003). Information aggregation with random ordering: Cascades and overconfidence. *The Economic Journal*, *113*(484), 166–189.

Pilditch, T. D. (2017). Opinion cascades and echo-chambers in online networks: A proof of concept agent-based model. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature neuroscience*, *19*(3), 366.

Schöbel, M., Rieskamp, J., & Huber, R. (2016). Social influences in sequential decision making. *PloS one*, *11*(1).

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, *99*(9), 5766–5771.

Whalen, A., Griffiths, T. L., & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive science*, *42*(1), 168–187.