Evolution and Human Behavior 40 (2019) 492–499

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Unleashing the BEAST: a brief measure of human social information use

Lucas Molleman[a,b,c,*], Ralf H.J.M. Kurvers[a], Wouter van den Bos[a,b,c,d]

[a] Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
[b] Amsterdam Brain and Cognition Center, University of Amsterdam, the Netherlands
[c] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands
[d] Max Planck University College London Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

ABSTRACT

Social information enables individuals to reduce uncertainty and increase decision accuracy across a broad range of domains. Intriguingly, individuals and populations consistently differ in social information use. Understanding the underlying causes of this variation has proven challenging due to the lack of a standardized paradigm to quantify social information use. Here we introduce the BEAST (Berlin Estimate AdjuStment Task); a brief (~5-min), simple, and incentive-compatible behavioural task to quantify individuals' propensities to use social information. In the task, participants observe an image with a number of animals and estimate the total number. Next, they receive another person's estimate, after which they provide a second estimate. An individual's average adjustment quantifies their propensity to use social information. We found that individuals' propensity to use social information is consistent within the task, has considerable test–retest reliability over 9 months, generalizes to other social learning tasks, and correlates with established self-reported measures of social conformity and social proximity. The BEAST thus reliably captures individual variation in social information use. We conclude by highlighting the BEAST's potential to serve as a flexible framework to assess the determinants of human social information use.

1. Introduction

Social information plays a central role in human behaviour. Observing others can help individuals make effective decisions and rapidly adjust to novel and changing environments (Boyd, Richerson, & Henrich, 2011; Hoppitt & Laland, 2013; Tomasello, 2009). Individuals' social information use (i.e., when and how the behaviour of others influences decision making) impacts social interactions and the performance of groups (Lorenz, Rauhut, Schweitzer, & Helbing, 2011; Torney, Lorenzi, Couzin, & Levin, 2015; Van den Berg, Molleman, & Weissing, 2015; Wolf, Kurvers, Ward, Krause, & Krause, 2013) and drives the spread of skills, knowledge, and social norms through populations (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Henrich, 2015; Whiten, Caldwell, & Mesoudi, 2016). Moreover, it directly affects key social and economic outcomes in domains ranging from voting (Bond et al., 2012) to consumption (Moretti, 2011) and health decisions (Centola, 2011).

While theoretical models have traditionally assumed that social information use does not differ between individuals (reviewed in Aoki & Feldman, 2014), recent empirical research has documented substantial variation in social information use between societies (Jayles et al., 2017; Mesoudi, Chang, Murray, & Lu, 2015; Molleman & Gächter, 2018), individuals (Efferson, Lalive, Richerson, McElreath, & Lubell, 2008; McElreath et al., 2005; Molleman, Van den Berg, & Weissing, 2014; Toelch, Bruce, Newson, Richerson, & Reader, 2014), and across the life span (Legare, 2017; Mesoudi, Chang, Dall, & Thornton, 2016; Morgan, Laland, & Harris, 2015; van Leeuwen et al., 2018). Despite its importance for human decision making across a range of domains, the causes of this variation in social information use remain poorly understood.

The use of social information is known to be affected by the external environment. For example, social information use has been shown to increase with task difficulty and environmental variability (Derex, Beugin, Godelle, & Raymond, 2013; Derex & Boyd, 2016; Efferson et al., 2008; McElreath et al., 2008, 2005; Mesoudi, 2008; Mesoudi et al., 2015; Molleman et al., 2014; Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012; Muthukrishna, Morgan, & Henrich, 2016; Toelch et al., 2014; Van den Berg et al., 2015). A few studies have explored potential links between individual characteristics and the use of social information. Muthukrishna et al. (2016) showed that individuals with higher IQ scores used less social information. Toelch et al. (2014) showed that individuals who self-reported as more collectivist than others used more social information. These studies, however, primarily used between-individual differences to correct for individual-specific responses to different environmental conditions, and these differences explained only a small proportion of variation in social information use. Furthermore, it is unclear to what extent the various paradigms used in the literature robustly identify individual differences in social information use, and whether their behavioural measures generalize to other decision-making settings. Although standardized tasks for quantifying

* Corresponding author at: Centre for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.
E-mail address: l.s.molleman@uva.nl (L. Molleman).

https://doi.org/10.1016/j.evolhumbehav.2019.06.005
Received 3 January 2019; Received in revised form 24 May 2019; Accepted 24 June 2019
1090-5138/ © 2019 Elsevier Inc. All rights reserved.

individual differences have been developed in other domains of decision making (e.g., risk taking and social preferences (Chuang & Schechter, 2015; Lejuez et al., 2002)), at present, there is no standardized behavioural paradigm to measure social information use.

We address this gap by introducing a standardized behavioural task measuring individuals' propensities to use social information: the BEAST (Berlin Estimate AdjuStment Task). We designed the task to be simple, brief, incentive-compatible, and flexible. Simplicity is important to make the task accessible to participants from different cultural and demographic backgrounds and with a wide range of cognitive capacities. Brevity facilitates the adoption of the task in large cohort studies. Incentive-compatibility enables researchers to measure actual behaviour (with monetary consequences) rather than relying on self-reports. Finally, flexibility accommodates a range of possible extensions to address more specific research questions (e.g., how is social information use affected by task difficulty, demonstrator identity, or group size?).

To establish whether the BEAST reliably measures individual differences in social information use, we conducted two sets of experiments examining three key aspects of its validity: (i) test–retest reliability (Experiment 1); (ii) convergence with social information use in other decision-making settings (Experiment 2); and (iii) correlations with established questionnaire-based constructs plausibly linked to social information use (Experiments 1 and 2).

The BEAST is a perceptual judgment task in which participants have to estimate how many animals are displayed in an image (Fig. 1). After entering their first estimate, participants receive social information (the estimate of another participant) and make a second estimate. We characterize an individual's propensity for social information use as their relative estimate adjustment towards the social information, averaged across five rounds of the task. Fig. 1 presents an overview of the BEAST, and Section 2.2 provides a detailed description.

We report on two experiments. We first examine the internal consistency of the BEAST, and then show that individuals' social information use (S) has considerable test–retest reliability (Experiment 1). Next, we assess whether S generalizes to other decision-making settings, demonstrating that the BEAST has considerable convergent validity (Experiment 2). Finally, we show that S correlates with self-reported social conformity and social proximity (Experiment 1 + 2).

2. Methods

2.1. General procedures

We recruited 324 participants (mean age: 35.3y, SD = 10.6; 57.5% male) from Amazon Mechanical Turk (MTurk), who completed the experiments through their web browsers. We restricted participation to MTurk workers from the United States and with a minimum approval rating of 90%. Participation was by informed consent, and the Ethics Committee of the Max Planck Institute for Human Development Berlin approved the studies (ref: ARC 2017/18). After completing the experimental tasks and questionnaires, participants received a unique code to receive their payment via MTurk. The experimental software was programmed in LIONESS Lab (Giamattei, Molleman, Seyed Yahosseini, & Gaechter, 2019; https://lioness-lab.org) and is available upon request from the corresponding author.

We conducted two experiments, each consisting of a set of behavioural tasks and questionnaires. We also recorded participants' age and gender. Each task started with instructions, followed by compulsory control questions to ensure participants' understanding. During the behavioural tasks, participants could earn points, which were converted into bonus earnings upon completion. The bonus came on top of a guaranteed participation fee.

Experiment 1 examined the temporal stability of the BEAST, by repeating the same experiment in three waves. In the first wave, we recruited 102 participants from MTurk (mean age ± SD = 35.7 ± 11.0; 54% male) to complete the BEAST (Section 2.2). We re-invited participants of Wave 1 two weeks (Wave 2) and nine months later (Wave 3), to complete the same task again. We followed

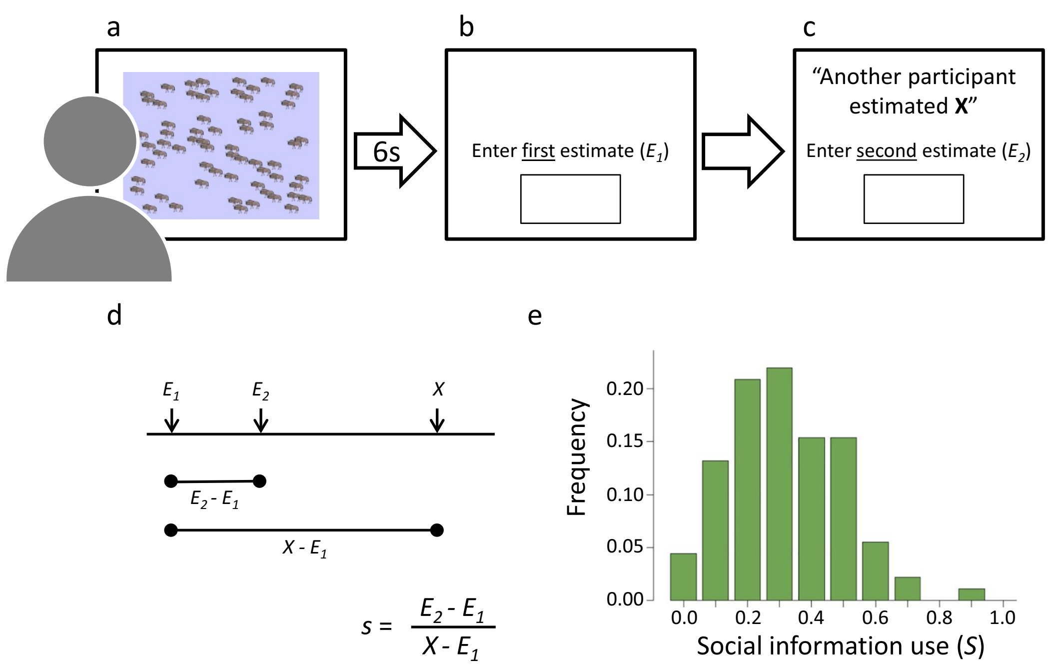

Fig. 1. Anatomy of the BEAST. a, Participants observe an image of a group of animals for 6 s and are asked to estimate the number of animals. b, Participants then enter their first estimate (E_1). c, Next, they observe social information (X; the first estimate of another participant who had already completed the task) and enter their second estimate (E_2). This procedure is repeated for five rounds. d, For each round, we define social information use (s) as the adjustment from E_1 to E_2 as a fraction of the distance between X and E_1. We characterize each individual's propensity to use social information as the mean value of s across the five rounds (denoted S). e, Distribution of S, illustrating that participants in our sample ranged from keeping their first estimate (i.e., S = 0) to moving 90% towards the social information (S = 0.9; N = 102 participants). See Fig. S1 for screenshots of the experiment; for a demo, see https://arc.mpib-berlin.mpg.de/BEAST.

the same procedures as in Wave 1 (e.g., participants observed the same images in the same order).

In each of the three waves, participants completed the BEAST and an unrelated sampling task (not reported here). On average, participants took about 10 min to complete the experimental sessions and earned $1.50 (including their participation fee), corresponding to an average hourly wage of $9.00. We invited all 100 participants who submitted a completion code to participate again in Wave 2 and 3 (in Wave 1, two of the 102 participants completed the task but did not submit a code). Participants received a message from MTurk stating that they 'qualified' for a follow-up experiment and that they would have two weeks to complete it. In Wave 2 (two weeks after Wave 1), 69 participants returned to complete the experiment; 47 returned in Wave 3 (nine months after Wave 1). Thirty-five participants completed all three waves.

Experiment 2 was designed to test whether social information use in the BEAST generalizes to other settings of social information use. To this end, we extended the set of tasks and questionnaires from Experiment 1 with the moving dots and bandit tasks (Sections 2.3 and 2.4). These tasks have been used to study social information use (Glowacki & Molleman, 2017; McElreath et al., 2005; Moussaïd, Herzog, Kämmer, & Hertwig, 2017; Newsome & Pare, 1988). The order of the three behavioural tasks was randomized. For Experiment 2, we recruited 222 participants from MTurk (mean age: 35.2, SD 10.4; 59% male). On average, participants took about 25 min to complete the experimental session and earned $4.25 (including participation fee), corresponding to an average hourly wage of $8.50.

2.2. Experimental task: the BEAST

The BEAST is a perceptual judgment task designed to measure individuals' social information use (Fig. 1). The task lasts for five rounds. In each round, participants observe an image with 50–100 animals (Fig. 1a; see Fig. S1 for screenshots). After 6 s, the image disappears and participants have to submit their estimate of the number of animals (Fig. 1b). The range of numbers of animals and the viewing time provides participants with a general impression of the total, but prevents them from actually counting the animals. The value the participants enter after viewing the image is their *first estimate* in a round (denoted E_1). We impose a time limit on submitting this estimate (30 s in the first round and 15 s thereafter) to avoid the possibility of participants taking a screenshot of the image and counting the animals at their leisure.

After submitting their first estimate (E_1), participants receive social information (X, the estimate of another person who already completed the same task; see below for details). Participants then make a *second estimate* (E_2; Fig. 1c). When making their second estimate, the screen shows the participant's first estimate and the social information alongside each other, as well as a timer that enforces participants to submit their second estimate within the time limit of 45 s. The BEAST operationalises social information use as the extent of participants' initial estimate adjustment (Fig. 1d). For each round, we calculate the relative distance a participant moves towards the social information as $s = (E_2 - E_1) / (X - E_1)$. Reordering this formula as $E_2 = (1 - s) \cdot E_1 + s \cdot X$ shows that the second estimate (E_2) in a round is a weighted average of an individual's initial estimate (E_1) and the social information (X), where s determines the relative weight put on social information (assuming that E_2 is a convex combination of E_1 and X over s).

We characterize an individual's propensity for social information use (S) as the average value of s across all five rounds. So, S is the central outcome measure of the BEAST. Prior to averaging, we excluded data from rounds in which a participant moved in the opposite direction of the social information ($s < 0$; 31 of 2200 cases in the full data set) or moved beyond the social information ($s > 1$; 57 cases). These adjustments outside the range of $0 \leq s \leq 1$ represent qualitatively different types of behaviour; that is, putting negative weight on social information (for $s < 0$) or not determining one's second estimate as a weighted average of the first estimate and social information (for $s > 1$). Including these cases (or treating them as the nearest value in the range [0,1]) in the analyses did not change any of the main results presented in this paper. Furthermore, we excluded eight cases in which a participant's first or second estimate was > 100 animals off the true value, assuming that these were due to typos in their estimates.

Participants do not receive feedback about their performance during the five rounds. This setup implies that participants could not learn about their own skill or the accuracy of the social information. Note that in many decision-making settings, such information might be available (e.g., in the bandit task; cf. Section 2.4). Performance feedback during the task potentially introduces substantial levels of noise into measurements of social information use as people dynamically update the value they ascribe to social information relative to other sources of information (Kurvers, Wolf, & Krause, 2014; Tump, Wolf, Krause, & Kurvers, 2018). After completing the five rounds (typically within 5 min), one of the 10 estimates was randomly selected for payment. When the selected estimate was correct, a participant received the maximum of 100 points (equalling $1.00). For each animal off the true value, 5 points were subtracted. Earnings could not fall below 0 points. Participants are thus incentivised to adjust their first estimate only when they think it will lead to an improvement in accuracy.

We controlled the distance between E_1 and X by selecting social information from a pre-recorded dataset of 100 MTurkers who completed the task without social information—a common procedure in advice-taking research (e.g., Yaniv, 2004). In particular, we selected social information 15–25% away from an individual's first estimate (see below for implementation details). Controlling the distance ensured that participants would have a relatively constant room for adjustment. We chose a moderate distance of social information because social information that is either very close or far from initial estimates tends to lead to little adjustment. (Hütter & Ache, 2016; Jayles et al., 2017; Moussaïd et al., 2017; Moussaïd, Kämmer, Analytis, & Neth, 2013; Schultze, Rakotoarisoa, & Schulz-Hard, 2015; Yaniv, 2004; Yaniv & Milyavsky, 2007).

In each round, we chose a specific piece of social information based on a participant's first estimate, using the true (correct) value as a reference point. In particular, we first calculated a 'target' value (X') of social information, using $X' = E_1 \cdot (1 + \Delta)$ if the first estimate was below the correct value, and $X' = E_1 \cdot (1 - \Delta)$ if the first estimate was above the correct value. The displayed value of X was the value in the pre-recorded sample closest to that 'target' value. When the first estimate was exactly correct, a random device determined whether the displayed social information was lower or higher than the first estimate. Social information could thus be lower or higher than a participant's first estimate. Furthermore, it was often (but not always) closer to the true value than a participant's first estimate. For rounds 1–5 of Experiment 1, Δ took the respective values of 0.25, 0.15, 0.20, 0.15, and 0.25. We collected the data for Experiment 2 in two sessions: In one session ($N = 99$) participants entered their estimates by filling out a numeric input box on their screen (as in Experiment 1); in the other ($N = 123$), participants entered their estimates with a slider. As regressions fitted to s in Experiment 1 revealed that the value of Δ had no systematic effect on the amount of adjustment, we kept Δ constant at 0.20 for the 'slider' session.

2.3. Moving dots task

The moving dots task is a well-studied perceptual estimation task in which participants have to estimate the main direction of 50 dots moving across their computer screen (Fig. 3a; Moussaïd et al., 2017; Newsome & Pare, 1988). The majority of dots move in random directions, but some move in a similar direction (the 'main direction'). This main direction is represented as an angle θ, taking a value between 0 and 360 degrees. We chose the parameters of the task such that, like in the BEAST, participants could get an overall impression of the true value while still being sufficiently uncertain of their own estimate. Based on Moussaïd et al. (2017), we chose 70% of the dots moving at random and 30% moving in the main direction, adding noise to the angle of each of these dots (avoiding identical parallel trajectories for dots moving in the main direction, which renders the task too easy). In particular, we added a value randomly chosen from a uniform distribution between −45 and 45 to the main direction θ (emulating the 'intermediate' difficulty level from Moussaïd et al. (2017)).

Participants completed five rounds of this task. In each round, after 10 s, the grid with the moving dots disappeared and participants had to estimate the main direction. After submitting their estimate, they were shown social information consisting of the estimate of another person. As in the BEAST, we had pre-recorded a sample of 100 MTurkers who completed the task without social information. For each round we chose social information by first calculating a 'target' value which was in the direction of the true value of θ, and differed from a participant by Δ degrees. The displayed social information was the value in the pre-recorded sample closest to that 'target' value. For rounds 1–5, Δ took the respective values of 25, 15, 20, 15, and 25. After observing the social information (displayed as a red arrow on their screen alongside their own first estimate, shown as a blue arrow; Fig. 3a), participants submitted a second estimate.

As in the BEAST, we measured a participant's social information use (M) as the distance between their second and first estimate, divided by the distance between the first estimate and the observed social information, averaged across the five rounds. Again, participants were rewarded for accuracy. At the end of the experimental session, one estimate was randomly selected for payment. If that estimate was exactly correct, participants would receive 100 points ($1.00). For each degree away from the true value, 1 point was subtracted, with the constraint that earnings could not become negative.

2.4. Bandit task

Two-armed bandit tasks, in which participants have to find out which of two options has a higher payoff, are commonly used to study individual and social learning (Fig. 3c; Efferson et al., 2008; Glowacki & Molleman, 2017; McElreath et al., 2008; Molleman et al., 2014). Our task was based on a seminal social learning study by McElreath et al. (2005), but adjusted it such that in each round, participants had to choose between individual or social information rather than having the option to gather social information in addition to their individual payoff information (Glowacki & Molleman, 2017).

In the task, participants repeatedly had to choose between two noisy options, one of which was associated with a higher expected payoff (38 points) than the other (30 points). We implemented noise by adding a stochastic term to the payoffs: an integer drawn—for each participant in each round separately—from a normal distribution with a mean of 0 and a SD of 12. This stochastic component was truncated such that actual payoffs in any given round were an integer value between 1 and 69 points (McElreath et al., 2005). This noise level implies that the option with the higher expected payoff yields higher payoffs in 67% of the cases (and a lower payoff in 31% of the cases, with the remaining 2% yielding the same payoff).

In each round (after the first), participants could collect information about the previous round. They could either observe the choices of three people who had previously completed the task (*social information*), or the payoff of their own choice (*individual information*). We operationalized social information use in the bandit task (B) as the number of rounds (out of 14) in which a participant chose social information (Glowacki & Molleman, 2017). As in the other tasks, participants' earnings depended on their performance. At the end of the task, the point earnings for all 15 rounds were added up and converted into money (500 points = $1.00). We used this conversion rate to keep the task close to the studies it was adapted from (Glowacki & Molleman, 2017; McElreath et al., 2008, 2005) while keeping the expected earnings in line with the BEAST and the moving dots task.

2.5. Questionnaires

At the end of Experiment 1 and 2, participants completed a set of questionnaires, aimed at measuring four constructs: social conformity, social proximity, individualism, and collectivism. (*i*) We measured social conformity using the Conformity Scale from Mehrabian and Stefl (1995) in which participants indicate their agreement with 11 statements (e.g., 'I often rely on, and act upon, the advice of others') on a 9-point scale. (*ii*) To measure social proximity, we used the 'Inclusion of the Other in the Self' scale (Aron, Aron, & Smollan, 1992), which has been validated for a wide range of classes of relationships (from close friends to distant relationships; Gächter, Starmer, & Tufano, 2015). In this task, participants are shown seven pairs of circles with varying degrees of overlap and asked to indicate which of the pairs best describes their relationship with others. In our implementation, these 'others' were the MTurkers whose estimates they could observe in the task. (*iii* + *iv*) In order to measure individualism and collectivism, participants indicated their agreement with a set of statements on a 9-point scale (Triandis & Gelfand, 1998). Both constructs had four associated statements (e.g., individualism: "I often do 'my own thing'"; collectivism: 'I feel good when I cooperate with others'). In all questionnaires, participants could leave items open if they preferred not to answer (for each of the reported correlations in Fig. 4, we omitted participants who left open an item for the respective questionnaire scale; out of our 324 participants, this procedure led to exclusion of 10 participants for the conformity scale, 10 for social proximity, 10 for individualism, and 11 for collectivism). We expected positive correlations between the BEAST's S and social conformity, social proximity, and collectivism, and a negative correlation with individualism.

2.6. Statistical analyses

All analyses were conducted in *R* v. 3.5.1 (R Core Team, 2015). Ordinary least squares (OLS) regression lines with confidence bands (Figs. 2, 3 and 4) were created using the package *visreg* (Breheny & Burchett, 2013). Repeatability was calculated using the *rptR* package (Stoffel, Nakagawa, & Schielzeth, 2017). Data and code associated with this paper can be found at https://github.com/LucasMolleman/Molleman_Kurvers_vandenBos_EHB2019_BEAST.

3. Results

3.1. General results and internal consistency

In Wave 1 of experiment 1, participants, on average, moved approx. 1/3 of the distance towards the observed social information when providing their second estimate (average $S = 0.318$; median $S = 0.298$; Fig. 1e), implying that people put about twice as much weight on their own estimate than on the estimates of others. The magnitude of this 'egocentric bias' is in line with earlier studies on social information use and advice taking (Bednarik & Schultze, 2015; Jayles et al., 2017; Madirolas & de Polavieja, 2015; Minson, Ross, & Liberman, 2011; Soll & Larrick, 2009; Soll & Mannes, 2011; Yaniv, 2004; Yaniv & Kleinberger, 2000).

Individuals' social information use, on average, ranged from keeping their own first estimate ($S = 0$) to moving 90% towards the observed social information ($S = 0.9$). The majority of participants put more weight on their own first estimate than on the observed social information. About 4% of participants did not adjust at all ($S = 0$). The majority of participants (82.4%) tended to move towards the social information, but did not move more than half the distance between their own first estimate and the observed social information ($0 < S \leq 0.5$). Only 13.7% of participants put more weight on the social information than on their own first estimate ($S > 0.5$). In individual rounds, adjustments were overwhelmingly in the expected range of $0 \leq s \leq 1$ (Fig. S2).

To test the internal consistency of the BEAST (that is, whether participants' social information use showed a consistent pattern over rounds) we calculated correlations in s across the five rounds of the task. We observed significant positive correlations for each pair of rounds (with Pearson's r ranging from 0.220 to 0.493; $P < .001$ for most pairwise correlations; max. $P = .033$; Table S1). In addition, a principle component analysis reveals that a single factor accounts for 52% of the variation across all rounds. Mixed model analyses further corroborated internal consistency, revealing that the random term 'individual' accounts for a substantial proportion of variation in (i) the likelihood of keeping one's own estimate versus moving towards the social information (Table S2), and (ii) the relative magnitude of the adjustment in cases where an adjustment was made (Table S3). Taken together, these results indicate that individuals tend to use social information in a consistent manner, and that individuals consistently differ from each other in their social information use.

On average, participants underestimated the number of animals by approx. 10.0% in their first estimate (Fig. S3). This underestimation is consistent with earlier studies on numerical estimation tasks (Izard & Dehaene, 2008; Jayles et al., 2017; Kao et al., 2018; Krueger, 1984). Importantly, individuals' mean accuracy in their first estimate did not affect their average adjustment S ($r = -0.002$, d.f. = 100, $t = -0.023$, $P = .982$; Fig. S4), suggesting that the propensity to use social information is independent of skill. Furthermore, social information use did not vary with age or gender (OLS regression fitted to individuals' S: $P > .291$ for both age and gender).

3.2. Social information use in the BEAST has substantial test–retest reliability

Individuals' social information use in both retests strongly correlated with their social information use in Wave 1 (Wave 2: Fig. 2a; Pearson's correlation coefficient $r = 0.587$; d.f. = 67, $t = 5.931$, $P < .001$; Wave 3: Fig. 2b; $r = 0.595$, d.f. = 45, $t = 4.971$, $P < .001$). These results indicate that the BEAST's behavioural measure S has substantial test–retest reliability.

3.3. The BEAST converges with other experimental measures of social information use

Social information use in the moving dots task correlated strongly and positively with social information use in the BEAST (Fig. 3b; $r = 0.480$, d.f. = 211; $t = 7.953$, $P < .001$). We observed a weak but significantly positive correlation between individuals' social information use in the bandit task and the BEAST (Fig. 3d; $r = 0.136$, d.f. = 219, $t = 2.026$, $P = .044$).

3.4. Social information use in the BEAST correlates with social conformity and proximity

To further gauge the convergent validity of the BEAST, we examined correlates of S with four individual psychological constructs plausibly related to social information use: social conformity, social proximity, collectivism, and individualism. S correlated positively with social conformity (Fig. 4a; $r = 0.346$, $t = 6.462$, d.f. = 308, $P < .001$), social proximity (Fig. 4b; $r = 0.241$, $t = 4.385$, d.f. = 311, $P < .001$), and, albeit weakly, with collectivism (Fig. 4c; $r = 0.118$, $t = 2.101$, d.f. = 310, $P = .036$). S did not correlate with individualism (Fig. 4d; $r = -0.032$, $t = -0.557$, d.f. = 311, $P = .578$). A regression analysis predicting S including all four measures detected significant effects only for social conformity and social proximity (Table S4).

4. Discussion

We have introduced and validated the BEAST, a brief behavioural task measuring people's propensity to use social information. We showed that individuals behave consistently across rounds of the BEAST, that individuals consistently differ from each other, and that the task's central measure S has substantial test–retest reliability over a 9-month period. Moreover, the correlations with social information use in different experimental paradigms and self-reported social conformity and proximity provide strong support for the convergent validity of S. In sum, the BEAST emerges as a robust and reliable measure for quantifying individual differences in social information use.

The BEAST provides a measure of individuals' social information use with substantial within-individual consistency (Tables S1, S2, and S3) and considerable time stability (Fig. 2). This suggests that the task measures a relatively invariant 'trait-like' form of behaviour. Individuals' performance in behavioural tasks often varies with their temporary state, such as motivation, mood, or fatigue level. This is one reason why behavioural measures tend to have lower levels of test–retest reliability than self-report measures (in which participants are often asked to describe themselves in general terms, or can base their responses on retrieving episodes of their own past behaviours from memory). Indeed, while self-report measures often have test-retest reliabilities of > 0.8 (e.g., Gnambs, 2014), behavioural tasks measuring impulsivity, risk taking, temporal discounting, and social preferences report test–retest correlations ranging from 0 to 0.7, with the majority on the lower end of this range (Chuang & Schechter, 2015; Frey, Pedroni, Mata, Rieskamp, & Hertwig, 2017; Pedroni et al., 2017; Weafer, Baggott, & de Wit, 2013). Comparing test–retest reliabilities between behavioural measures is not straightforward, as these measures can differ in many respects (e.g., methodology; number of observations, number of items or behavioural decisions underlying the measure; amount of individual variation measured by the task; time lag between test and retest; cognitive abilities of the sample population;

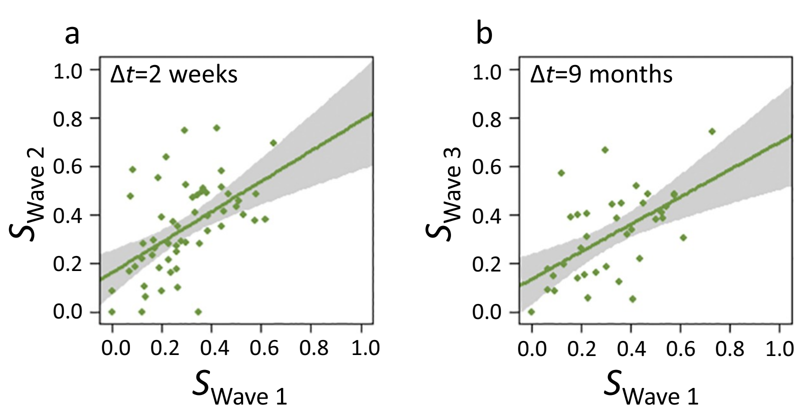

Fig. 2. Test–retest reliability of social information use in the BEAST. a, Social information use (S) of participants in Wave 1 positively correlated with their social information use two weeks later (Wave 2). b, Likewise, social information use of participants in Wave 1 positively correlated with their social information use nine months later (Wave 3). Each dot represents an individual; lines with 95% confidence intervals indicate the prediction of a linear OLS regression model.

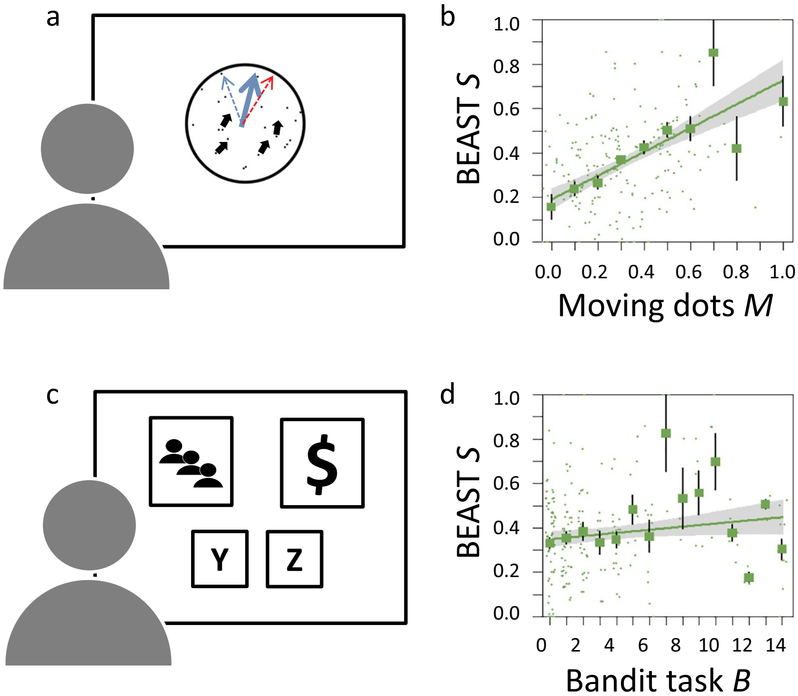

Fig. 3. Social information use in the BEAST generalizes to other settings. a, Moving dots task. In each of five rounds, participants observed 50 dots moving across their screen. After 10 s, the dots disappeared and participants estimated the main direction of the dots (thin blue arrow). Next, the estimate of another person was displayed (social information; thin red arrow) and participants entered a second estimate (thick blue arrow). b, There was a strong correlation between individuals' social information use in the BEAST (*S*) and the moving dots task (*M*). Squares with bars indicate the mean ± 1 S.E. for cohorts of *M*, where values of *M* were rounded to the nearest multiple of 0.1. c, Bandit task. For 15 rounds, participants could choose between two options with noisy payoffs (here shown as Y or Z; actual screens showed abstract shapes (Glowacki & Molleman, 2017)). In each round except the first, participants could collect information about the previous round. They could either observe the choices of three people who had already completed the task (social information; tile showing people), or the payoff of their own choice (individual information; tile showing a dollar sign). d, There was a weak, positive correlation between individuals' social information use in the BEAST (*S*) and the bandit task (*B*). Data points are horizontally jittered. In b and d, each dot represents an individual, and lines with 95% confidence intervals indicate the prediction of linear OLS regressions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

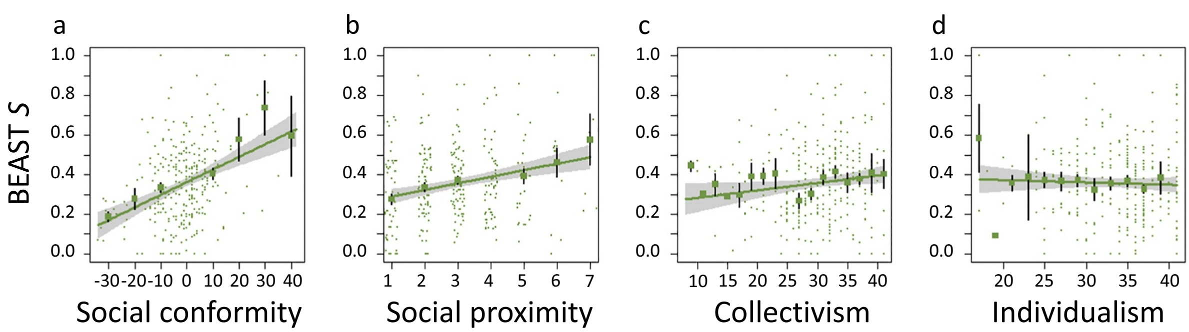

Fig. 4. Convergent validity: correlations of the BEAST with established self-report measures. a, Social conformity. b, Social proximity. c, Collectivism. d, Individualism. Each dot represents an individual, squares and bars show per-cohort mean ± 1 S.E., and lines and shaded bands show linear regressions lines and 95% confidence intervals.

and the domain of decision making). That said, the BEAST's test–retest correlation of ~0.6 after nine months (based on 5 data points per individual per wave), appears to compare favourably to behavioural measures from other domains.

On average, participants' first estimates were approx. 10.0% lower than the actual number of animals (Fig. S3). The distribution of deviations of first estimates from the true value shown in Figure S3 suggests that the current task settings (e.g., viewing time and number of animals) give participants a fairly accurate impression of the true value. At the same time, participants were not completely sure of their first estimates, since the majority of participants tended to adjust their first estimates. Importantly, participants' initial accuracy did not affect their average adjustment (Fig. S4), indicating that the BEAST's central measure *S* is not confounded by individuals' ability. We cannot rule out, however, that participants' confidence or their perception of their own skill level affected their social information use. A low level of social information use, for example, might reflect a high level of confidence in one's own first estimate. This seems particularly likely when estimates can be based on pre-existing knowledge (e.g., about trivia, historical dates, or caloric content of food; e.g., Jayles et al., 2017; Yaniv, 2004; Yaniv & Milyavsky, 2007). Such knowledge likely introduces systematic between-individual and between-item variation in confidence and, as a result, may bias measurements of an individual's propensity to use social information. Perceptual decision-making tasks like the BEAST do not involve such knowledge, thereby avoiding these potential confounds. Future research could evaluate whether—and if so, how—social information use in the BEAST is mediated by factors like confidence in one's own first estimates, a preference for consistency, or a desire to maintain one's self-image.

The consistency in social information use across decision-making settings (Fig. 3) and positive correlations with self-reported social conformity and proximity (Fig. 4) provide support for the BEAST's convergent validity. Individuals' social information use in the BEAST and the moving dots task strongly correlated (Fig. 3b). This is perhaps unsurprising given that both tasks are perceptual estimation tasks with a single piece of social information and the possibility of financial reward. It is perhaps more surprising that the BEAST's *S* also correlates (albeit weakly) with the inclination to rely on social information in the bandit task (Fig. 3d). Gathering social information and responding to it are different components of social information use, and our results suggest that these two components are not independent. Future research could assess the predictive value of the BEAST for behaviour in other domains where social information plays a key role, such as cooperation (Van den Berg et al., 2015), rule-following (Gächter & Schulz, 2016), and compliance to social norms (Bicchieri, 2005), as well as its predictive validity for people's social information use outside of the controlled environment of a decision-making experiment.

We designed the BEAST to be intuitive and simple. Participants from our MTurk sample readily completed the task. Alongside the convenience sample from MTurk, we administered the BEAST in subject pools of varying backgrounds and cognitive abilities. These include Colombian indigenous fishermen and farmers, German and Indian school children, and British adolescents with conduct problems. We will report the results of these studies in separate papers (in preparation). In all samples tested thus far, participants rapidly and easily comprehended the instructions and completed the task within a few minutes. Across all these samples, individuals were overwhelmingly (> 95%) in the expected range of $0 \leq s \leq 1$. This indicates that the BEAST is suitable for samples from a wide range of demographics, making it a promising tool for understanding variation in social information use across individuals, cultures, and the life span.

With its simple design and easily manipulable parameters, the BEAST can serve as a versatile framework for examining key assumptions and predictions of theories of social information use (Boyd & Richerson, 1985; Hoppitt & Laland, 2013; Kendal et al., 2018). Examples include (*i*) to test how social information use is impacted by properties of the external environment. That is, whether, and to what extent,social information use increases with task difficulty (e.g., reflected in the number of animals displayed and the viewing time), the number of observable peers, and the extent to which peer estimates are in agreement with each other; (*ii*) to investigate how individuals might or might not flexibly shift their social information use when exposed to different types ofexternal environments (as stated in *i*); (*iii*) to delineate how social information use is shaped by characteristics of the decision maker, the observed peer(s), and their relationship (e.g., in terms of their performance, reputation or dominance rank, in terms of social proximity, or with the peer being a parent, teacher, or stranger); (*iv*) to provide a systematic and quantitative assessment of how the use of social information depends on its distance to a participant's own estimate, and (*v*) to test how people learn to use social information over time as they gain more experience with and knowledge about the accuracy of social information relative to their own personal estimates.

To conclude, we believe that the BEAST can be a valuable complement to researchers' data-collection process. When time and resources are limited, researchers can readily append this brief task to their experimental sessions. A widespread adoption of the BEAST for collecting data in different participant samples would make it possible to explore how social information use is shaped by people's local (social) ecology, socio-demographics, conditions during development, cognitive capacities, and personality characteristics. Ultimately, this would facilitate a comprehensive understanding of the key determinants of people's social information use.

Competing interests

The authors declare no competing interests.

Data availability

The data for this paper can be accessed via https://github.com/ [https://github.com/LucasMolleman/Molleman_Kurvers_vandenBos_EHB2019_BEAST].

Acknowledgements

We thank the members of the Center for Adaptive Rationality at the Max Planck Institute for Human Development and the attendants of the Bonn–Berlin–Leipzig meeting in Bonn (March 2018) for useful comments. We are grateful to Mr. Herbert Krause for providing the images of animals. We thank Andrea Gradassi for help with programming and data collection, and Deb Ain for proofreading the manuscript. L.M. and W.v.d.B. were supported by Open Research Area (ID 176); L.M. is further supported by an Amsterdam Brain and Cognition Project grant 2018. W.v.d.B. is further supported by the Jacobs Foundation, European Research Council grant (ERC-2018-StG-803338) and the Netherlands Organization for Scientific Research grant (NWO-VIDI 016.Vidi.185.068).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2019.06.005.

References

Aoki, K., & Feldman, M. W. (2014). Evolution of learning strategies in temporally and spatially variable environments: A review of theory. *Theoretical Population Biology, 91*, 3–19. https://doi.org/10.1016/j.tpb.2013.10.004.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Bednarik, P., & Schultze, T. (2015). The effectiveness of imperfect weighting in advice taking. *Judgment and Decision making, 10*, 265–276.

van den Berg, P., Molleman, L., & Weissing, F. J. (2015). Focus on the success of others leads to selfish behavior. *Proceedings of the National Academy of Sciences, 112*(9), 2912–2917. https://doi.org/10.1073/pnas.1417203112.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

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.

Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process.* University of Chicago Press.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences, 108*, 10918–10925.

Breheny, P., & Burchett, W. (2013). Visualization of regression models using visreg. *R Package,* 1–15.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach.(MPB-16). Vol. 16.* Princeton University press.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science, 334*(6060), 1269–1272.

Chuang, Y., & Schechter, L. (2015). Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of Development Economics, 117*, 151–170.

Core Team, R. (2015). R: A Language and Environment for Statistical Computing. Retrieved from http://www.R-project.org/.

Derex, M., Beugin, M.-P., Godelle, B., & Raymond, M. (2013). Experimental evidence for the influence of group size on cultural complexity. *Nature, 503*(7476), 389–391. https://doi.org/10.1038/nature12774.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences, 113*(11), 2982–2987.

Efferson, C., Lalive, R., Richerson, P. J., McElreath, R., & Lubell, M. (2008). Conformists and mavericks: The empirics of frequency-dependent cultural transmission. *Evolution and Human Behavior, 29*(1), 56–64.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances, 3*(10), https://doi.org/10.1126/sciadv.1701381.

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature, 531*, 496–499.

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the closeness of relationships: A comprehensive evaluation of the "Inclusion of the Other in the Self" scale. *PLoS One, 10*(6), e0129478. https://doi.org/10.1371/journal.pone.0129478.

Giamattei, M., Molleman, L., Seyed Yahosseini, K., & Gaechter, S. (2019). *LIONESS Lab – a free web-based platform for conducting interactive experiments online.* Available at SSRN https://ssrn.com/abstract=3329384https://doi.org/10.2139/ssrn.3329384. SSRN scholarly paper no. ID 3329384). Retrieved from Social Science Research network website https://papers.ssrn.com/abstract=3329384.

Glowacki, L., & Molleman, L. (2017). Subsistence styles shape human social learning strategies. *Nature Human Behaviour, 1*, 0098.

Gnambs, T. (2014). A meta-analysis of dependability coefficients (test–retest reliabilities) for measures of the big five. *Journal of Research in Personality, 52*, 20–28. https://doi.org/10.1016/j.jrp.2014.06.003.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton University Press.

Hoppitt, W., & Laland, K. N. (2013). *Social learning: An introduction to mechanisms, methods, and models.* Princeton University Press.

Hütter, M., & Ache, F. (2016). Seeking advice: A sampling approach to advice taking. *Judgment & Decision Making, 11*(4), 401.

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition, 106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004.

Jayles, B., Kim, H., Escobedo, R., Cezera, S., Blanchet, A., Kameda, T., ... Theraulaz, G. (2017). How social information can improve estimation accuracy in human groups. *Proceedings of the National Academy of Sciences,* 201703695.

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., ... Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of the Royal Society Interface, 15*(141), 20180130. https://doi.org/10.1098/rsif.2018.0130.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences, 22*(7), 651–665. https://doi.org/10.1016/j.tics.2018.04.003.

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics, 35*(6), 536–542. https://doi.org/10.3758/BF03205949.

Kurvers, R. H. J. M., Wolf, M., & Krause, J. (2014). Humans use social information to adjust their quorum thresholds adaptively in a simulated predator detection experiment. *Behavioral Ecology and Sociobiology, 68*(3), 449–456. https://doi.org/10.1007/s00265-013-1659-6.

van Leeuwen, E. J. C., Cohen, E., Collier-Baker, E., Rapold, C. J., Schäfer, M., Schütte, S., & Haun, D. B. M. (2018). The development of human social learning across seven societies. *Nature Communications, 9*(1), 2076. https://doi.org/10.1038/s41467-018-04468-2.

Legare, C. H. (2017). Cumulative cultural learning: Development and diversity. *Proceedings of the National Academy of Sciences, 114*(30), 7877–7883. https://doi.org/10.1073/pnas.1620743114.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75–84. https://doi.org/10.1037/1076-898X.8.2.75.

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. https://doi.org/10.1073/pnas.1008636108.

Madirolas, G., & De Polavieja, G. G. (2015). Improving collective estimations using resistance to social influence. *PLoS Computational Biology, 11*(11), e1004594. https://doi.org/10.1371/journal.pcbi.1004594.

McElreath, R., Bell, A. V., Efferson, C., Lubell, M., Richerson, P. J., & Waring, T. (2008). Beyond existence and aiming outside the laboratory: Estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1509), 3515–3528.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., ... Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior, 26*(6), 483–508.

Mehrabian, A., & Stefl, C. A. (1995). Basic temperament components of loneliness, shyness, and conformity. *Social Behavior and Personality: An International Journal, 23*(3), 253–263.

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: Adaptive landscapes, producer–scrounger dynamics, and informational access costs. *Evolution and Human Behavior, 29*(5), 350–363.

Mesoudi, A., Chang, L., Dall, S. R., & Thornton, A. (2016). The evolution of individual and cultural variation in social learning. *Trends in Ecology & Evolution, 31*, 215–225.

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 of London B: Biological Sciences, 282*(1798), 20142209. https://doi.org/10.1098/rspb.2014.2209.

Minson, J., Ross, V., & Liberman, L. (2011). *Two to Tango: Effects of Collaboration and Disagreement on Dyadic Judgment.* Retrieved September 10, 2018, fromhttps://doi.org/10.1177/0146167211410436.

Molleman, L., & Gächter, S. (2018). Societal background influences social learning in cooperative decision making. *Evolution and Human Behavior, 39*(5), 547–555. https://doi.org/10.1016/j.evolhumbehav.2018.05.007.

Molleman, L., van den Berg, P., & Weissing, F. J. (2014). *Consistent individual differences in human social learning strategies*.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies, 78*(1), 356–393.

Morgan, T. J. H., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: Effects of uncertainty and consensus. *Developmental Science, 18*(4), 511–524. https://doi.org/10.1111/desc.12231.

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.

Moussaïd, M., Herzog, S. M., Kämmer, J. E., & Hertwig, R. (2017). Reach and speed of judgment propagation in the laboratory. *Proceedings of the National Academy of Sciences, 114*, 201611998.

Moussaïd, M., Kämmer, J. E., Analytis, P. P., & Neth, H. (2013). Social influence and the collective dynamics of opinion formation. *PLoS One, 8*(11), e78433.

Muthukrishna, M., Morgan, T. J., & Henrich, J. (2016). The when and who of social learning and conformist transmission. *Evolution and Human Behavior, 37*(1), 10–20.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience, 8*(6), 2201–2211. https://doi.org/10.1523/JNEUROSCI.08-06-02201.1988.

Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour, 1*(11), 803–809. https://doi.org/10.1038/s41562-017-0219-x.

Schultze, T., Rakotoarisoa, A.-F., & Schulz-Hard, S. (2015). Effects of distance between initial estimates and advice on advice utilization. *Judgment and Decision making, 10*(2), 144–172.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 780–805. https://doi.org/10.1037/a0015145.

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting, 27*(1), 81–102. https://doi.org/10.1016/j.ijforecast.2010.05.003.

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution, 8*(11), 1639–1644.

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.

Tomasello, M. (2009). *The cultural origins of human cognition.* Harvard University Press.

Torney, C. J., Lorenzi, T., Couzin, I. D., & Levin, S. A. (2015). Social information use and the evolution of unresponsiveness in collective systems. *Journal of the Royal Society Interface, 12*(103), 20140893. https://doi.org/10.1098/rsif.2014.0893.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74*(1), 118.

Tump, A. N., Wolf, M., Krause, J., & Kurvers, R. H. J. M. (2018). Individuals fail to reap the collective benefits of diversity because of over-reliance on personal information. *Journal of the Royal Society Interface, 15*(142), 20180155. https://doi.org/10.1098/rsif.2018.0155.

Weafer, J., Baggott, M. J., & de Wit, H. (2013). Test–retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. *Experimental and Clinical Psychopharmacology, 21*(6), 475–481. https://doi.org/10.1037/a0033659.

Whiten, A., Caldwell, C. A., & Mesoudi, A. (2016). Cultural diffusion in humans and other animals. *Current Opinion in Psychology, 8*, 15–21. https://doi.org/10.1016/j.copsyc.2015.09.002.

Wolf, M., Kurvers, R. H. J. M., Ward, A. J. W., Krause, S., & Krause, J. (2013). Accurate decisions in an uncertain world: Collective cognition increases true positives while decreasing false positives. *Proceedings of the Royal Society of London B: Biological Sciences, 280*(1756), 20122777. https://doi.org/10.1098/rspb.2012.2777.

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93*(1), 1–13. https://doi.org/10.1016/j.obhdp.2003.08.002.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes, 83*(2), 260–281. https://doi.org/10.1006/obhd.2000.2909.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes, 103*(1), 104–120.