

## OBSERVATION

# Together, Slowly but Surely: The Role of Social Interaction and Feedback on the Build-Up of Benefit in Collective Decision-Making

Bahador Bahrami  
University College London and Aarhus University

Karsten Olsen, Dan Bang, and Andreas Roepstorff  
Aarhus University

Geraint Rees  
University College London

Chris Frith  
University College London and Aarhus University

That objective reference is necessary for formation of reliable beliefs about the external world is almost axiomatic. However, Condorcet (1785) suggested that purely subjective information—if shared and combined via social interaction—is enough for accurate understanding of the external world. We asked if social interaction and objective reference contribute differently to the formation and build-up of collective perceptual beliefs. In three experiments, dyads made individual and collective perceptual decisions in a two-interval, forced-choice, visual search task. In Experiment 1, participants negotiated their collective decisions with each other verbally and received feedback about accuracy at the end of each trial. In Experiment 2, feedback was not given. In Experiment 3, communication was not allowed but feedback was provided. Social interaction (Experiments 1 and 2 vs. 3) resulted in a significant collective benefit in perceptual decisions. When feedback was not available a collective benefit was not initially obtained but emerged through practice to the extent that in the second half of the experiments, collective benefits obtained with (Experiment 1) and without (Experiment 2) feedback were robust and statistically indistinguishable. Taken together, this work demonstrates that social interaction was necessary for build-up of reliable collaborative benefit, whereas objective reference only accelerated the process but—given enough opportunity for practice—was not necessary for building up successful cooperation.

*Keywords:* collective decision-making, feedback, 2-AFC, social interaction

In the story of “The Elephant in the Dark”, the medieval Farsi-speaking poet Rumi masterfully portrayed the limitations places on beliefs by noisy sensory perception (Tourage, 2007). Late one evening, an Indian circus arrived at a village. The more curious villagers sneaked into the elephant’s stable. In absolute darkness, they made observations by touching the elephant’s body. When they returned to their families, their accounts, constrained by their limited sensory experiences, gave widely divergent images of the elephant. Rumi concluded that “light” —an external source

of objective reference—is necessary for formation of reliable beliefs about the external world. Without objective reference, beliefs will be purely subjective.

Here we test this notion empirically against the hypothesis that, if individual observers without objective reference, such as Rumi’s villagers, have the opportunity to share their experiences through repeated social interaction, they could ultimately achieve as accurate beliefs as with objective reference. We borrow this approach from the Marquis de Condorcet (1785), an Enlightenment mathe-

---

This article was published Online First October 10, 2011.

Bahador Bahrami, UCL Institute of Cognitive Neuroscience, University College London, Centre of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark; Karsten Olsen, Dan Bang, and Andreas Roepstorff, Interacting Minds Project, Institute of Anthropology, Archaeology, Linguistics, Aarhus University, and Centre of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark; Geraint Rees, UCL Institute of Cognitive Neuroscience and Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, England, United Kingdom; Chris Frith, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, England, United Kingdom, and Interacting Minds Project, Institute of Anthropology, Archaeology, Linguistics, Aarhus University,

and Centre of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark.

This work was supported by a British Academy postdoctoral fellowship (to Bahador Bahrami), the Danish National Research Foundation and the Danish Research Council for Culture and Communication, the European Union MindBridge project (to Bahador Bahrami, Andreas Roepstorff, and Chris Frith), and by the Wellcome Trust (to Geraint Rees). Support from the MINDLab UNIK initiative at Aarhus University was funded by the Danish Ministry of Science, Technology and Innovation.

Correspondence concerning this article should be addressed to Bahador Bahrami, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, United Kingdom. E-mail: bbahrami@gmail.com

matician and political philosopher, who proposed that, by sharing and combining purely subjective information via social interaction, humans could construct reliable beliefs about the external world.

Previous research on collective decision making (Henry, Strickland, Yorges, & Ladd, 1996) suggests that access to feedback (i.e., objective reference) enables groups to evaluate the reliability of individual members more effectively. However, even with feedback, groups of interacting individuals seems unable to exceed their best individuals in tasks involving general knowledge, organizational management skills (Tindale, 1989), and jury decision tasks (Hastie & Kameda, 2005). However, recent evidence suggests that interacting groups can achieve a reliable collective benefit over and above their best individuals in a perceptual task (Bahrami et al., 2010), even in the absence of feedback, thus supporting Condorcet's view.

Collective endeavor needs coordination and mutual understanding, which often takes time, effort, and experience to establish. Previous research reviewed above has placed little emphasis on the role of experience and learning in cooperation. Here we tested the

role of social interaction and feedback on the build-up of effective collective perceptual decision making.

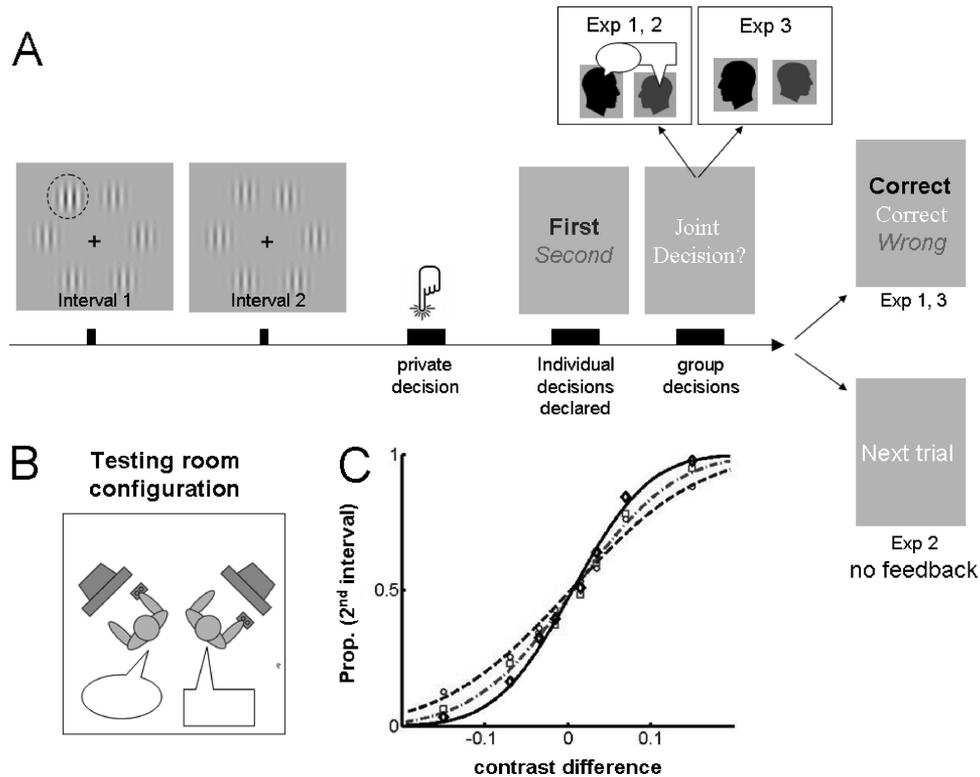
## Methods

### Participants

Seventy-two adult healthy male participants with normal or corrected-to-normal vision were recruited in Aarhus, Denmark, plus an additional eight participants in London, U.K. (Experiment 1 [ $N = 22$ ], mean age  $\pm$  standard deviation [ $SD$ ]:  $28.30 \pm 6.27$ ; Experiment 2 [ $N = 30$ ], mean age  $\pm$   $SD$ :  $26.30 \pm 5.3$ ; and Experiment 3 [ $N = 28$ ]), mean age  $\pm$   $SD$ :  $27.30 \pm 6.1$ ). Members of each dyad knew each other. No participant was recruited for more than one experiment. All experiments were approved by the local ethics committee.

### Stimulus, Task Design, and Procedure

Sitting in the same room, each participant faced his own screen (Figure 1B), placed at right angles to the other. The two displays



**Figure 1.** (A) Sequential schema of events in Experiments 1–3. Each trial started with two stimulus intervals. Visual stimuli consisted of six vertically oriented Gabor patches that were displayed equidistantly around an imaginary circle. One (randomly selected) interval contained the target (here indicated by the dashed circle) of higher contrast. Participants then indicated their individual decisions privately. If they disagreed, in Experiments 1 and 2 participants negotiated a joint decision; in Experiment 3, one (randomly assigned) participant made a joint decision on behalf of the group without any interaction. Feedback about accuracy was provided in Experiments 1 and 3 but not in Experiment 2. (B) Testing setup. (C) Group average psychometric functions relating the individual and group choice to stimulus strength in Experiment 1. The X axis shows the contrast difference at the oddball location (i.e., contrast in the second interval minus contrast in the first). The Y axis shows the proportion of trials in which the target was reported to be in the second interval. Circles: average performance of the less sensitive dyad members; squares: average performance of the more sensitive dyad members; diamonds: average performance of the dyads. The curves are the best fit to a cumulative Gaussian function.

(800 × 600 resolution, Fujitsu Siemens AMILO SL 3220W, 22", viewing distance ~ 57 cm, output luminance linearized, background luminance 62.5 Cd/m<sup>2</sup>) were connected to the same computer controlled by Cogent (<http://www.vislab.ucl.ac.uk/cogent.php>) for MATLAB (Mathworks Inc). Within each experimental session, one participant responded with the keyboard and the other with the mouse.

In every trial (Figure 1A), participants viewed a visual search array containing a contrast-defined target plus five distracters in a two-interval forced-choice design. Each trial was initiated by one participant. After an initial central fixation cross (width: 0.75 degrees; random duration, range 500–1000 ms), the two stimulus intervals (duration: 85 ms; separated by a blank screen—duration: 1000 ms) ensued. The stimulus set (Figure 1A) comprised six vertically oriented Gabor patches (standard deviation of the Gaussian envelope: 0.45 degrees; spatial frequency: 1.5 cycles/degree; nontarget contrast: 10%) equally spaced on a circle (eccentricity: 8 degrees). Target contrast was 1.5%, 3.5%, 7.0%, or 15% higher than the nontarget pedestal. Target location, contrast, and interval were randomized across trials. A central question mark was displayed after the second interval and remained until both participants had responded.

Participants first made private individual decisions about the target interval, which were then displayed publicly (Figure 1A) using color codes to distinguish keyboard (blue) and mouse (yellow) responses. If private decisions disagreed, a joint decision was requested. The keyboard participant announced the joint decision in odd trials; the mouse participant in even trials. In Experiments 1–2, participants negotiated this joint decision through communication. In Experiment 3, one of the two participants arbitrated for the dyad without communicating with the other; participants were instructed not to talk, wore earphones to eliminate auditory communication via unintentional utterances and were separated by a cardboard screen. The experimenter was present throughout testing to ensure instructions were observed.

In Experiments 1 and 3, participants received feedback about outcomes (“CORRECT” or “WRONG”), one for each participant (keyboard: blue; mouse: yellow) and one for the dyad (white).

After one practice block of 16 trials, two main experimental sessions were conducted. Each main session consisted of eight blocks of 16 trials. Participants switched places (and thereby screen and response device) at the end of Session 1.

## Analysis

To quantify perceptual sensitivity, the slope of the psychometric function relating target contrast to choice was estimated (Figure 1C) for individuals (Figure 1C, dashed lines) and dyad (Figure 1C, solid line) separately. A cumulative Gaussian function with parameters bias,  $b$ , and variance,  $\sigma^2$  was fitted to each obtained psychometric function by a probit regression model employing the *glmfit* function in MATLAB (Mathworks Inc). A participant with bias  $b$  and variance  $\sigma^2$  would have a psychometric curve, denoted  $P(\Delta c)$  where  $\Delta c$  is the contrast difference between the second and first presentations, given by

$$P(\Delta c) = H\left(\frac{\Delta c + b}{\sigma}\right),$$

where  $H(z)$  is the cumulative normal function,

$$H(z) \equiv \int_{-\infty}^z \frac{dt}{(2\pi)^{1/2}} \exp[-t^2/2].$$

Here, the psychometric curve,  $P(\Delta c)$ , corresponds to the probability of saying that target was in the second interval. Given this definitions for  $P(\Delta c)$ , the variance is related to the maximum slope of the psychometric curve, denoted  $s$ , via

$$s = \frac{1}{(2\pi\sigma^2)^{1/2}}.$$

This slope parameter quantifies the perceptual sensitivity. The collective benefit over and above the individual dyad members was defined as the ratio of the dyad slope to that of the more sensitive dyad member.

## Sliding Window Analysis

To investigate the development of collective-decision making, we performed a sliding window analysis (see Figure 3). For each dyad, the first window of selection sampled the Trials 1–80. Psychometric functions were estimated for the individuals and dyad and collective benefit was determined for the sampled trials. The window was then slid by one step to sample the Trials 8–87. This process was repeated until the sliding window reached the last 80 trials (176–256). We used this approach rather than splitting the data into small but nonoverlapping blocks because constructing and estimating reliable psychometric functions requires a large number of trials.

## Results

### Qualitative Observations

Participants in Experiment 1–2 were encouraged to discuss and try any useful strategy. Some groups split the task spatially (e.g., upper and lower half) or temporally (first and second intervals). However, such strategies were never sustained. Participants invariably preferred attending to both intervals in all locations.

Dyad conversations focused on participants' confidence in their decisions. However, only one group among the 30 interacting dyads (Experiments 1–2) spontaneously used an explicit numerical scale to express and compare confidence. All other groups used more everyday expressions such as “I was not so sure” and “I saw it clearly.” On a trial-by-trial basis, participants tended to align to each other's confidence expressions (e.g., if one person initiated with “I did not see anything,” their partner was likely to respond with an expression involving “see”). In general, the content of conversations tended to diminish with practice. Participants converged to a small set of expressions toward the end of the experiment. A more detailed linguistic analysis of these aspects of the interactive confidence sharing process is work in progress (Fusaroli et al., submitted).

### Quantitative Results

Social interaction enabled the groups to exceed their best performing members. A mixed  $3 \times 2$  analysis of variance

(ANOVA) on sensitivity (i.e., the slope of the psychometric function) with experiment as between-subject factor with three levels and decision maker as within-subject factor with two levels (better individual vs. dyad, Figure 2) showed a significant main effect for decision maker,  $F(1, 41) = 26.16$ ;  $p < 0.0001$ , but not for experiment,  $F(2, 41) = 1.8$ ;  $p > 0.17$ . Importantly, there was a significant interaction between decision maker and experiment,  $F(2, 41) = 7.38$ ;  $p < .002$ . Post hoc comparisons showed that dyads outperformed the more sensitive member of the group in Experiments 1,  $t(14) = 5.2$ ;  $p < .0001$ ; paired  $t$  test and 2  $t(14) = 3.85$ ,  $p < .002$ ; paired  $t$  test, but not in Experiment 3,  $t(14) = 0.01$ ;  $p > .9$ ; paired  $t$  test. These data showed clearly that social interaction was necessary for achieving a collective benefit and that this benefit was statistically identical with and without feedback (i.e., objective reference). They do not, however, identify the role of social interaction and feedback in the build-up of collective perceptual decision making.

The sliding window analysis (see Figure 3) indicated that with feedback, collective benefit was fairly stable across time both when communication was (Experiment 1, black squares) and was not (Experiment 3, dark gray triangles) allowed. Without feedback, a different pattern was observed (Experiment 2, light gray circles): little collective benefit was initially achieved in the first 1/3 of the experiment. However, a rising trend emerged leading to a consistent collective benefit in the latter 2/3 of the experiment. In the final third of the trials, interactive groups achieved similar levels of collective benefit irrespective of feedback.

In order to statistically qualify these observations, we employed a mixed  $3 \times 2$  ANOVA (experiment: between-group factor with three levels; session: within-group factor with two levels; see Figure 4) with collective benefit as the dependent variable. This showed a significant main effect of session,  $F(1, 41) = 6.75$ ,  $p = .013$  and experiment,  $F(2, 41) = 7.20$ ,  $p = .002$ , and a significant interaction between session and experi-

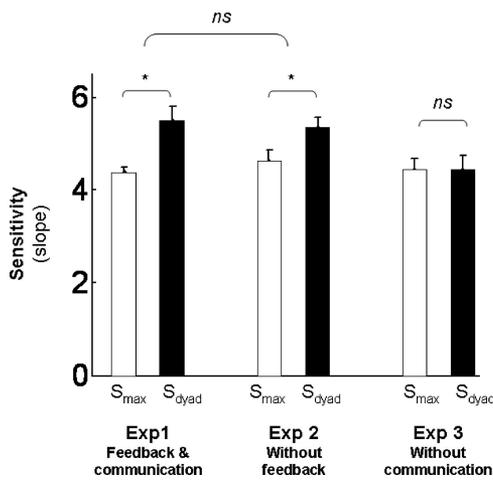


Figure 2. Average sensitivity (i.e., the slope of the psychometric function—see Figure 1C) is plotted for the best performing member of each group ( $S_{max}$ ) and the dyads ( $S_{dyad}$ ). Interacting dyads (Experiments 1 and 2) exceeded their own best performing individual. Without interaction, providing feedback (Experiment 3) did not afford any group benefit.

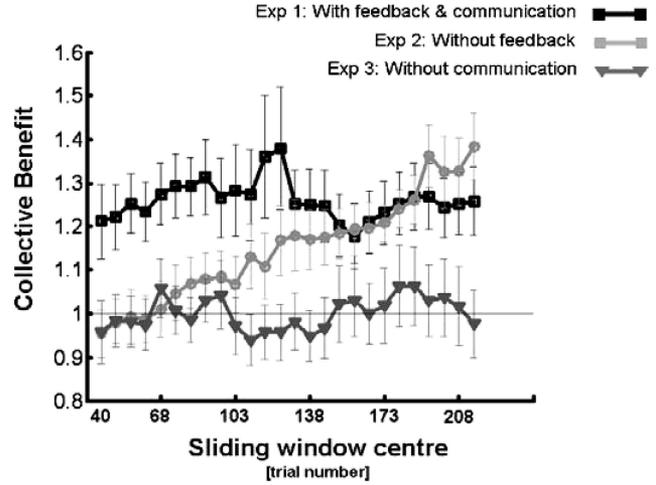


Figure 3. Sliding window analysis shows that collective benefit was gradually accumulated in the absence of feedback. Collective benefit (Y axis) is plotted against the trial number at the center of the sampling window. Error bars represent one standard error of the mean.

ment,  $F(2, 41) = 4.98$ ,  $p = .012$ . With feedback (Figure 4, black line), dyads achieved consistent and significant collective benefit in both experimental sessions, (one-sample  $t$  test comparing each data point to the horizontal “no benefit” line  $y = 1$ , for Session 1,  $t(14) = 2.72$ ,  $p = .016$ ; for Session 2,  $t(14) = 5.29$ ,  $p < .001$ ), with no significant difference between sessions, (paired  $t$  test,  $t(14) = 1.13$ ,  $p > .25$ ). Without feedback (Figure 4, gray dashed line), significant collective benefit was only achieved in the second session, (one-sample  $t$  test, for Session 1,  $t(14) = 0.7$ ,  $p > .5$ ; for Session 2,  $t(14) = 5.6$ ,  $p < .001$ ), and a remarkable impact of social interaction over time was demonstrated by a highly significant difference between the two sessions, paired  $t$  test,  $t(14) = 4.5$ ,  $p < .001$ . Finally, without social interaction (Figure 4; gray dotted line), access to feedback did not produce any benefit in either session ( $p > .6$  for both sessions; one-sample  $t$  test comparing to 1) and there was no evidence

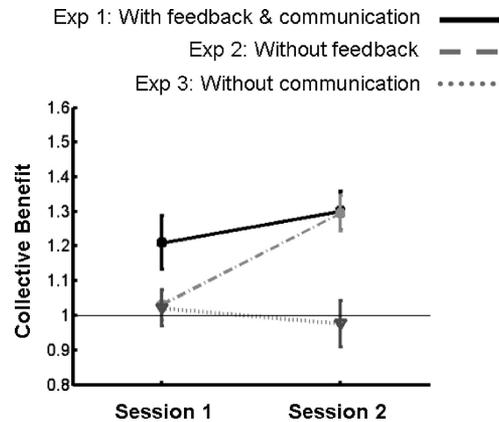


Figure 4. Average collective benefit accrued from joint decision making is plotted for the two sessions of the experiments. Error bars represent one standard error of the mean.

for any improvement in collective benefit over time ( $p > .5$  paired  $t$  test).

### Discussion

Our results showed that social interaction resulted in significant collective benefit in perceptual decisions. When feedback was not available, collective benefit was not initially obtained but emerged through practice, so that in the second half of the experimental run, collective benefits obtained with and without feedback were statistically indistinguishable. With feedback, communicating dyads achieved a robust collective benefit from the beginning of the first session. Thus, the role of objective reference may be to accelerate the process of belief formation at the group level.

Our results reject Rumi's notion that objective reference is indispensable for forming reliable beliefs about the external world. Instead, they support Condorcet's (1785) view that sharing and combining purely subjective information via social interaction is sufficient for forming reliable beliefs—that is to say, sufficient given adequate opportunity for practice.

Perceptual decision making in isolated individuals is enhanced by feedback (Herzog & Fahle, 1997; Herzog & Fahle, 1999; Seitz, Kim, & Watanabe, 2009), even when feedback is motivationally relevant but informationally void (Shibata, Yamagishi, Ishii, & Kawato, 2009). Similar but smaller improvements are also observed without feedback (Ball & Sekuler, 1987; Poggio, Fahle, & Edelman, 1992) indicating that feedback is not always necessary for perceptual learning. Feedback also plays a crucial role in nonperceptual collective decision-making. As discussed earlier, several studies have shown that feedback is necessary for improving collective decision accuracy (Henry, Strickland, Yorges, & Ladd, 1996; Tindale, 1989; Hastie & Kameda, 2005). Our work goes beyond this to examine the impact of social interaction and feedback on the development of collective perceptual decision making. Our findings raise a number of issues for future research.

What are the computational characteristics of learning in collective belief formation? One computational model (Bahrami et al., 2010) proposes that collective beliefs are constructed via “confidence sharing,” where confidence is the observer's subjective estimate of probability of being correct on a particular trial. Importantly, this model does not include any role for social learning and instead assumes that the impact of practice is negligible. Although the data presented here show that this assumption is reasonable when feedback is provided, this model is not consistent with the findings from the “no feedback” experiment. Future research should seek to explain the construction of effective collective behavior over the course of repeated interactions in the absence of feedback. Several recent computational models have been proposed for social interaction based on principles of associative reinforcement learning (Behrens, Hunt, Woolrich, & Rushworth, 2008; Behrens, Hunt, & Rushworth, 2009; Hampton, Bossaerts, & O'Doherty, 2008). But it is difficult to see how these models can account for what seems to be unsupervised social learning (i.e., without any explicit reinforcement or feedback).

One appealing conjecture is that exchanging subjective confidence (individuals' metacognitive awareness of their perceptual decisions), could replace feedback and reinforce social learning. In

this view, metacognition provides an imperfect and noisy but informative estimate of the true state of the world (i.e., decision outcomes) when the outcome is not readily available. The learning process therefore takes longer to develop without feedback but given enough practice, leads to as much collective benefit as with feedback. As such, we speculate that a functional role of metacognitive awareness may be to replace missing reinforcement when decision outcomes are not available. Communication and confidence sharing afford access to two, rather than one reinforcement signal eventually leading to collective benefit.

Metacognitive accuracy varies across individuals and such variability has been correlated with gray matter volume in anterior medial prefrontal cortex (Fleming, Weil, Nagy, Dolan, & Rees, 2010). Our conjecture thus predicts that social collective learning without feedback should be quicker and more successful in individuals with better metacognition and thus with larger gray matter volume in anterior medial prefrontal cortex.

### References

- Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., & Frith, C. D. (2010). Optimally interacting minds. *Science*, *329*, 1081–1085. doi:10.1126/science.1185718
- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*, 953–965. doi:10.1016/0042-6989(87)90011-3
- Behrens, T. E., Hunt, L. T., & Rushworth, M. F. (2009). The computation of social behavior. *Science*, *324*, 1160–1164. doi:10.1126/science.1169694
- Behrens, T. E., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. (2008). Associative learning of social value. *Nature*, *456*, 245–249. doi:10.1038/nature07538
- Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris, France: De l'Imprimerie Royale.
- Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*, 1541–1543. doi:10.1126/science.1191883
- Fusaroli, R., Bahrami, B. P., Olsen, K., Roepstorff, A., Frith, C., & Tylen, K. (2010). Unfolding cognitive coordination in a state of uncertainty. In: *Language as Social Coordination*. Manuscript submitted for publication.
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*, 6741–6746. doi:10.1073/pnas.0711099105
- Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review*, *112*, 494–508. doi:10.1037/0033-295X.112.2.494
- Henry, R. A., Strickland, O. J., Yorges, S. L., & Ladd, D. (1996). Helping groups determine their most accurate member: The role of outcome feedback. *Journal of Applied Social Psychology*, *26*, 1153–1170. doi:10.1111/j.1559-1816.1996.tb02290.x
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*, 2133–2141. doi:10.1016/S0042-6989(97)00043-6
- Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, *39*, 4232–4243. doi:10.1016/S0042-6989(99)00138-8
- Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*, 1018–1021. doi:10.1126/science.1589770
- Seitz, A. R., Kim, D., & Watanabe, T. (2009). Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron*, *61*, 700–707. doi:10.1016/j.neuron.2009.01.016

Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision Research*, *49*, 2574–2585. doi:10.1016/j.visres.2009.06.009

Tindale, R. S. (1989). Group vs individual information-processing—the effects of outcome feedback on decision-making. *Organizational Behavior and Human Decision Processes*, *44*, 454–473. doi:10.1016/0749-5978(89)90019-8

Tourage, M. (2007). *Rumi and the hermeneutics of eroticism*. Leiden, The Netherlands: BRILL. doi:10.1163/ej.9789004163539.i-260

Received April 19, 2011  
 Revision received July 25, 2011  
 Accepted July 27, 2011 ■

### Correction to Potter et al. (2011)

The article “Attention Blinks for Selection, Not Perception or Memory: Reading Sentences and Reporting Targets,” by Mary C. Potter, Brad Wyble, and Jennifer Olejarczyk (*Journal of Experimental Psychology: Human Perception and Performance*, 2011, Vol. 37, No. 6, pp. 1915–1923) contained several production-related errors.

In Table 1, the critical words should have been shown in bold for Experiment 1 only. A corrected table appears below.

Table 1  
*Examples of a Sentence in Each Experiment*

---

Experiment 1: Our tabby cat <b>chased</b> <sup>a</sup> a <b>mouse</b> all around the backyard
Experiment 2: Our tabby cat CHASED a MOUSE all around the backyard
Experiment 3: Our 6666 tabby cat 2222 chased a mouse all around the backyard
Experiment 4: Our six tabby cat two chased a mouse all around the backyard

---

*Note.* Just one lag condition is illustrated; see text for the lag and serial positions in each experiment.

<sup>a</sup> The critical words in Experiment 1 were red but in the same font as the other words; here they are shown in bold.

DOI: 10.1037/a0026992