




Benefits of spontaneous confidence alignment between dyad members



Collective Intelligence
Volume 1:2: 1–14
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/26339137221126915
journals.sagepub.com/home/col


Niccolò Pescetelli 

Department of Experimental Psychology, University of Oxford, Oxford, UK

Collective Intelligence Lab, Department of Humanities and Social Sciences, New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ 07102, USA

PSi, SURU Together Ltd, The Black Church, St. Mary's Place, Dublin D07P4AX, Ireland

Nick Yeung 

Department of Experimental Psychology, University of Oxford, Oxford, UK

Abstract

In many domains, imitating others' behaviour can help individuals to solve problems that would be too difficult or too complex for the individuals. In collective decision making tasks, people have been shown to use confidence as a means to communicate the uncertainty surrounding internal noisy estimates. Here, we show that confidence alignment, namely, shifting average confidence between dyad members towards each other, naturally emerges when interacting with others' opinions. This alignment has a measurable impact on group performance as well as the accuracy of individual members following information exchange. It is suggested that confidence alignment arises among individuals from the necessity of minimising confidence variation arising from task-unrelated variables (trait confidence), while at the same time maximising variation arising from stimulus characteristics (state confidence).

Keywords

confidence alignment, confidence, metacognition, decision-making, dyads

Date received: 13 February 2022; revised: 24 July 2022; accepted: 29 August 2022

Corresponding author:

Niccolò Pescetelli, Collective Intelligence Lab, Department of Humanities and Social Sciences, New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ 07102, USA.

Email: niccolo.pescetelli@njit.edu



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Significance Statement

Behavioral alignment between members of a social group are at the heart of many dynamical properties in collective intelligence. In humans, individuals aligning beliefs with social partners can lead to positive effects such as social learning and innovation transmission, or group losses such as social contagion, herding and groupthink. The authors experimentally study alignment of decision confidence in dyads. Decision confidence is related to several effects in group opinion dynamics, including belief escalation, polarization and optimal information integration in joint decision-making. However, it remained unknown whether phenomena of confidence alignment emerge spontaneously or as a strategic adaptation to reach higher group performance. The authors experimentally demonstrate that confidence alignment spontaneously emerges from interaction with social partners and remains aligned after the interaction period. The results suggest that belief alignment is likely to take place even in the absence of collective incentives and suggest interventions in several applied fields. For example, on social media platforms, belief alignment and overconfidence often galvanize problems of misinformation. Interventions introducing low-confidence virtual agents might diffuse belief escalation and increase calibration. Another significant application is to the treatment of clinical populations with metacognitive deficits. The results suggest that possible treatments could pair patients with virtual tutors showing complementary confidence calibration patterns compared to the target patient. Under-confident patients might benefit from interaction with over-confident tutors and vice versa.

Introduction

Individual behaviours spread from one individual to others via observation, interaction or simple exposure to other people (Frith, 2007; Zajonc et al., 1987; Wheeler, 1966). Peer imitation is thought to be an important component of social as well as cultural learning (Heyes, 2017; Tomasello, 2009). By imitating, either voluntarily or automatically, social agents synchronise behaviour (Gallup et al., 2012; Milgram et al., 1969) and emotions (Dezecache et al., 2013; Kramer et al., 2014). In many species this allows individual organisms to magnify perceptual and cognitive capabilities at minimal costs (Bonabeau et al., 1999; Kennedy et al., 2001; Handegard et al., 2012). For instance, fireflies synchronise their flashing behaviour to increase female fireflies' recognition of suitable mates (Moiseff and Copeland, 2010), and fish schools distribute the sensing of their environment beyond what could be possible by a single fish (Hein et al., 2015). Here, we explore the importance of behavioral alignment in human decision-making, specifically as it relates to communication of uncertainty and confidence in choices made.

Decisions in joint tasks have been a recent focus of investigation, with studies showing that collective performance is crucially dependent on individual members sharing their confidence in their individual answers (Migdal et al., 2012; Sorkin et al., 2001; Bahrami et al., 2010). In particular, dyads of individuals solving a task together have been characterized as near-Bayesian optimal information integrators, who weight each member's private answers by their relative uncertainty (the inverse of expressed confidence) to determine the collective decision (Bahrami et al., 2010). Under ideal conditions, this strategy can create a

wisdom of crowds benefit whereby the group as a whole outperforms even its best performing member (Bahrami et al., 2010, 2012; Valeriani et al., 2017). However, the collective benefit depends on groups members' confidence reports being both calibrated (such that, across decisions, an individual is objectively more likely to be correct when they are more confident) and aligned (such that, across members of the group making a decision, more confident individuals are more likely to be correct), and the group adopting the decision of the most confident individual. These ideal conditions are not always met. For example, confidence is often mismatched with objective accuracy, particularly when compared across individuals (Kruger and Dunning, 1999). Moreover, groups may not integrate opinions optimally. For example, when one individual in a dyad is clearly better at the task than the other, the optimal strategy is to always choose the best individual's response, but people in these situations have been shown to suffer an equality bias (Mahmoodi et al., 2015) whereby, notwithstanding obvious accuracy differences, they prefer to adopt an equal opinion-weighting policy. More broadly, average calibration between group members has been shown to positively scale with a dyad's benefit over its members' separate performance (Pescetelli et al., 2016), indicating that accurately reporting one's own internal states is crucial for achieving good collective outcomes.

It follows that an important potential constraint on effective group decision making is mis-alignment of individual members' confidence reports, such that a group is swayed towards less optimal but more confidently expressed opinions. Experimental studies have confirmed the everyday intuition that some people are systematically more confident than others regardless of objective performance

(Ais et al., 2016), and such ‘trait’ confidence has been shown to be influenced by socio-economic variables including gender (Barber and Odean, 2001), profession (Broihanne et al., 2014), mental-health (Huq et al., 1988), personality (Campbell et al., 2004) and culture (Mann, 1998). But confidence in social tasks is important only insofar it carries task-relevant information, or according to a Bayesian interpretation, as long as it scales with the inverse-variance of the internal representation of external task-relevant variables (Pouget et al., 2016; Meyniel et al., 2015). In this respect, confidence is known to reliably track important aspects of stimulus information, like volatility, variance, task difficulty and internal perceptual noise (Kiani et al., 2014; Zylberberg et al., 2016; Yeung and Summerfield, 2012). Thus, in order to optimally share information among group members working together, individuals should minimise variability in their ‘trait’ confidence (task-irrelevant), while at the same time preserving variability in ‘state’ confidence (task-relevant). Early results on advice-taking behaviour (Yates et al., 1996) showed that over- or under-confidence do not hurt the reputation of an advisor as long as confidence judgments are predictive of actual outcomes – that is, high confidence *resolution* (Fleming and Lau, 2014). The result suggests that the informativeness of confidence judgments rather than absolute confidence matters for social communication.

Indeed it seems that people are willing and able to align their expressions of confidence – for example, shifting individual members’ average confidence toward a central mean when their objective accuracy is similar – in order to solve this coordination problem (Bang et al., 2017). In joint perceptual tasks, the presence of confidence alignment in verbal communication has been reported, suggesting that,

when participants adopted common linguistic expressions of confidence, dyadic performance exceeded best performing members (Fusaroli et al., 2012). On the contrary, indiscriminate alignment, that is, alignment of linguistic components that were not strictly task-relevant did not predict any group benefit. Meanwhile, in the context of a cooperative task in which the more confident dyad member’s decision is automatically adopted as the group’s final choice, alignment has been observed as a correlation between dyadic pairs’ mean confidence (Bang et al., 2017), leading to the idea that people apply a ‘private-to-public mapping’ of their internal confidence to an expressed report, with the latter sensitive to social context (in particular, the average confidence of the virtual partner in the task).

The present study extended these studies to ask whether this form of alignment in group decision-making can be observed even in the absence of any explicit requirement for cooperation, as has been observed with other forms of peer imitation (Frith, 2007; Zajonc et al., 1987; Wheeler, 1966; Kramer et al., 2014). We ask further whether this alignment has benefits for the individual as well as the group, over what time-course it develops and whether it persists beyond the period of interaction. Contrary to existing studies demonstrating alignment in cooperative tasks (Bang et al., 2017), we show that confidence alignment spontaneously emerges also in tasks that do not require cooperation. In our experiment (Figure 1), we record people’s judgements and decision confidence in a series of binary perceptual decisions (Boldt and Yeung, 2015) which are performed in parallel with a partner participant. We compare confidence distribution alignment across three phases of the experiment. During phases I and III, people did not see their

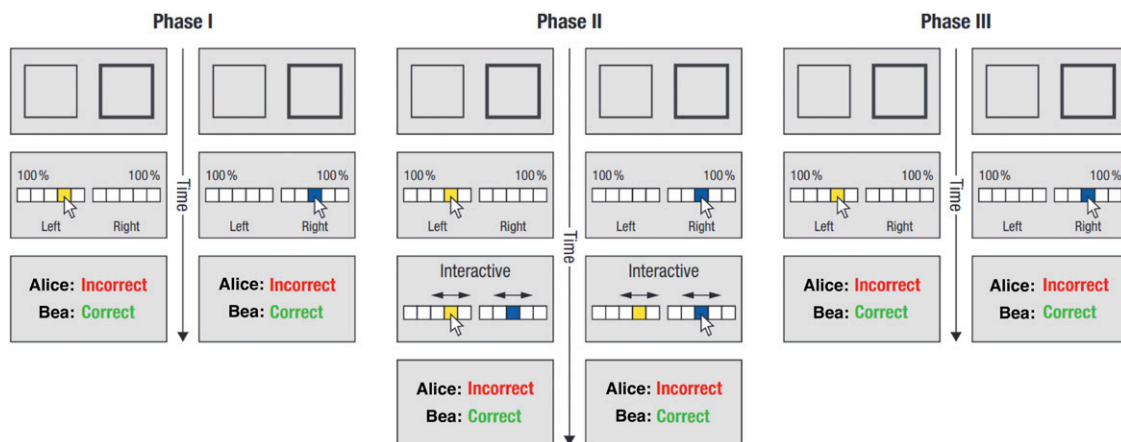


Figure 1. Experimental paradigm employed. During phase I, participants perform in parallel the perceptual task expressing on each trial a choice and a confidence judgment. In phase II of the experiment, after expressing their own private views, participants are shown their own and their partner’s moment-by-moment choice and confidence and are incentivised to continuously and accurately update their initial responses. In phase III, participants are again asked to perform the task alone and their choice and confidence are recorded on every trial.

partners' confidence, but only received feedback about their own and their partner's decision accuracy at the end of each trial. In phase II, after providing their own private judgments, people could dynamically interact with their partner via a slider interface to update their original decision in light of their partner's decision and associated confidence (which were also updated, visibly, in real-time), but still each participant made separate final decisions on each trial (and were rewarded only according to their own performance). Of interest was whether this interaction resulted in confidence alignment, over what timescale and whether alignment was beneficial.

The present study aims at showing in a highly controlled social environment (a) whether confidence alignment can be induced by manipulating the interaction between individuals sharing information about a common perceptual task and (b) test whether confidence alignment is beneficial for group performance.

Method

Participants

Participants (mean age: 21.14 ± 2.70) were 24 same-gender dyads (19 female dyads). Having same-gender dyads avoided confounds due to potential gender differences in the use of confidence scales and it represents standard practice in the literature (Buchan et al., 2008; Mahmoodi et al., 2015). Volunteers were recruited through local advertisements websites and University recruitment platform. Interested volunteers were asked to bring a friend along on the day of the experiment. Both participants in each dyad were compensated for their time and according to performance with money and/or university credits.

Task and manipulation

Participants sat on the two sides of a desk divided by a wooden occluder, each given a separate LCD monitor, keyboard and mouse. All devices were controlled by the same computer (Dell OptiPlex 9020). The experiment comprised 432 trials divided in three phases of six blocks each. All trials involved a perceptual decision task of judging whether a left- or right-hand box contained more dots (Boldt and Yeung, 2015), with each participant first registering their independent binary judgement (left/right) and associated confidence, without any interaction. In phases I and III, after both participants had registered their answers, feedback was given about both participant's accuracy and a new trial began. During both of these phases, participants could not see their partner's decision, confidence or cursor. Performance for each participant was separately titrated to 70% using a 2-down-1-up staircase function (Levitt, 1971). In phase II of the experiment, after

both participants had confirmed their individual answers, a 4-s social exchange part began, where each dyad member was informed about their partner's real-time choice and confidence, which was presented on the same response scale as the participant's own response cursor. The social part started with each member's cursor in the same position as indicated during the individual part of the trial. During the social part, confidence changes were recorded continuously in time, by incentivising participants to accurately report their moment-by-moment confidence (see [Supplementary Information](#) for details), while the cursors' positions along the confidence scale were continuously recorded at 5 Hz (21 data points per trial). For the whole duration of the social part, if a member updated their confidence, this would instantly appear also on their partner's scale and vice versa. This led to a situation where participants were not only informed of their partner's original opinions but also how those opinions were changing in real-time as a function of their own updates (Figure 1) (Pescetelli and Yeung, 2020). Feedback was provided at the end of this interaction and then a new trial began.

The correct answer (i.e. whether the left or right box had more dots) was randomized across trials and was always the same for both members of a dyad on each trial. However, unbeknownst to participants, the actual stimulus differed so as to match all participants' overall performance. Specifically, the two boxes contained dots arranged in random order on a 20×20 grid, with one box containing $200 + d$ dots and the other containing $200 - d$ dots. Perceptual difficulty was titrated continuously through the task to $\sim 70\%$ correct response rate for each participant in parallel, by manipulating the d parameter with a 2-down 1-up procedure (Levitt, 1971): Difficulty was increased by a unit amount after every two correct responses and decreased by the same amount after every error. Thus, within each dyad, the two individuals would see stimuli that differed in the specific configuration of dots, so that their accuracy could be matched even if their ability differed. Responses about choice and confidence judgment were recorded at once using a semi-continuous confidence scale ranging from 100% *Sure Left* to 100% *Sure Right*, with the middle level removed to force participants commit to one or other decision. Each side of the scale comprised 50 levels. Participants entered their responses by clicking with their mouse along the scale and confirmed their answer by pressing space bar. On top of the time compensation, an extra bonus could be achieved by accurately reporting final confidence during the individual part and continuous confidence during the social part. Before the beginning of phase I, one block of six trials served as practice with the perceptual task. Before the beginning of phase II, participants received a new set of instructions explaining the new input modalities and incentive scheme during the social exchange part, but no practice was given.

Results

Basic task performance

Reflecting our use of a psychophysical 2-down-1-up staircase procedure, perceptual decision accuracy was relatively stable across phases of the experiment, with mean accuracy of participants' separate private-phase decisions being 70.3%, 71.0% and 71.3% in phases I, II and III, respectively. Reflecting the benefits of social information sharing as has been observed previously (Pescetelli and Yeung, 2021), accuracy was higher after dyadic interaction than before it in phase II (73.6% vs. 71.0%, $t(47) = 5.95$, $p < .0001$). Reflecting the typically-observed individual differences in subjective confidence, even when objective performance is closely matched (range 69.2%–72.0%), mean confidence ranged from 60.8%–84.1% across participants. Confidence calibration, defined as the area under the type II ROC curve (Fleming and Lau, 2014), was above chance ($M = 0.61$, $SD = 0.04$, range=[0.52–0.69], $t(47) = 19.36$, $p < .001$).

Confidence alignment

The central test of the experiment was a comparison of confidence alignment over the three experimental phases. As noted, mean confidence varied substantially across participants (full confidence distributions for each participant and phase are shown in Supplementary Information). Of particular interest was whether confidence differences between dyad members decreased during interaction in phase II. For this, we focus on confidence as it was separately registered by the two dyad members during the private, pre-interaction phase on each trial (i.e. before they had a chance to interact, see their partner's potentially disagreeing opinion, and revise their decision and confidence accordingly). Empirical Pearson's correlation coefficients between these pre-advice confidence means of participants belonging to same dyad is shown in Figure 2 top panel. The observed correlation was weakly positive in phase I and not statistically significant ($\rho = 0.14$, $p = .51$), sharply increased during the interaction phase of the experiment where a significant correlation was observed ($\rho = 0.41$, $p = .045$), suggesting a causal relation with the experimental manipulation, and then decreased somewhat but not back to initial levels in phase III when participants again performed the task without interaction ($\rho = 0.33$, $p = .11$). These results were not strongly altered when partialling out variations in accuracy and threshold (dark grey bars in Figure 2 top panel). We used a bootstrapping procedure with 1000 permutations per experimental phase to compare empirical correlation of confidence means between dyad members with the distribution of confidence correlations that would be expected by randomly pairing participants.

Results show that, compared to a bootstrapped distribution, the observed correlation between confidence means went from the 74 percentile in phase I to 97 percentile in phase II and 94 percentile in phase III (Figure 3).

To apply formal statistical comparison across phases, the absolute difference in mean confidence between the two dyad members was taken as a dyad-level measure of alignment. Alignment reflects the fact that participants shifted their pre-advice confidence distributions toward each other, so that values closer to zero indicate greater alignment. Figure 2, bottom panel, plots distance as negative values, to correspond with correlation coefficients plotted in the top panel, where more positive values indicate closer alignment. Our intervention (phase II) produced a ~36% reduction in mean confidence difference. Values were then subjected to a one-way within-subjects ANOVA which revealed that phase significantly altered alignment ($F(2, 46) = 4.48$, $p = .01$, $\eta_G^2 = .06$). Planned comparisons showed that overlap was significantly larger during the interaction phase (phase II) than during the pre-interaction phase (phase I, $t(23) = 2.80$, $p = .01$, $d = 0.58$), confirming that alignment was directly caused by the experimental manipulation. The difference in mean confidence was numerically smaller in phase III than phase I but the difference was not statistically significant ($t(23) = 1.76$, $p = .09$, $d = 0.35$), and nor was that between phase II and phase III ($t(23) = 1.22$, $p = .23$, $d = 0.25$).

To explore the time-course of confidence alignment, Figure 4 plots the negative sliding-window mean (window size = 10 trials) of the trial-by-trial decision confidence difference between the two members of each dyad

$$alignment = -movAvg_{10}(|C_{S1} - C_{S2}|) \quad (1)$$

Alignment of mean confidence is high at the start of phase I because, before the psychophysical staircase settles for each participant, all participants tend to have low confidence. Alignment then tends to reduce throughout phase I as participants settle into their idiosyncratic patterns of reporting confidence in their decisions (Ais et al., 2016). There is then a clear effect of the interaction phase (phase II). Confidence alignment, so calculated, showed a steep increase at the start of phase II that trends upwards throughout the period of interaction and then partially reduces during phase III when interaction ceases.

The next set of analyses investigated what are the drivers of alignment and whether confidence shifts are symmetrical between the two participants. Noting that task difficulty was set to match the accuracy of all participants at ~ 70%, two potential factors were considered, namely average confidence and confidence calibration during the

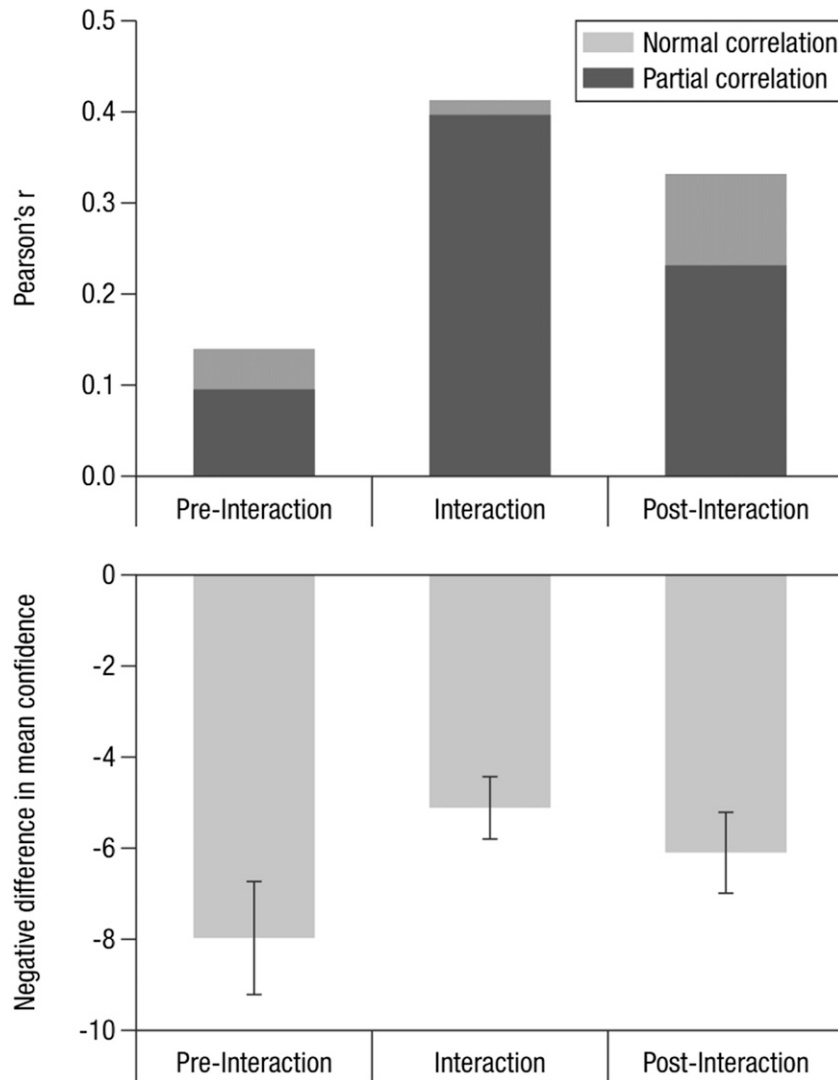


Figure 2. Change in confidence alignment between participants of similar dyads over the course of three experimental phases. (a): Pearson's correlation coefficient between means of the pre-social exchange confidence distributions of members belonging to the same dyads (light grey bars); similar results were obtained when controlling for variations in accuracy and perceptual threshold (dark grey bars). (b): negative distance between confidence means of pre-social exchange distributions.

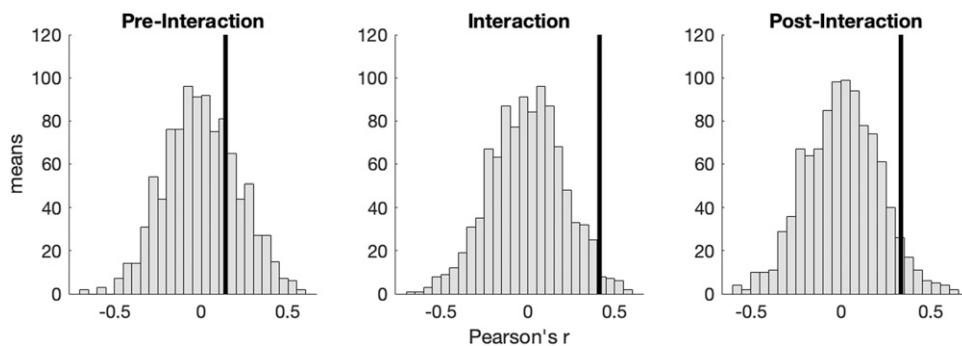


Figure 3. Bootstrap procedure across the three experimental phases. Correlation coefficients between means of the empirically observed confidence distributions are compared against the distribution of correlation coefficients expected by a random pairing procedure.

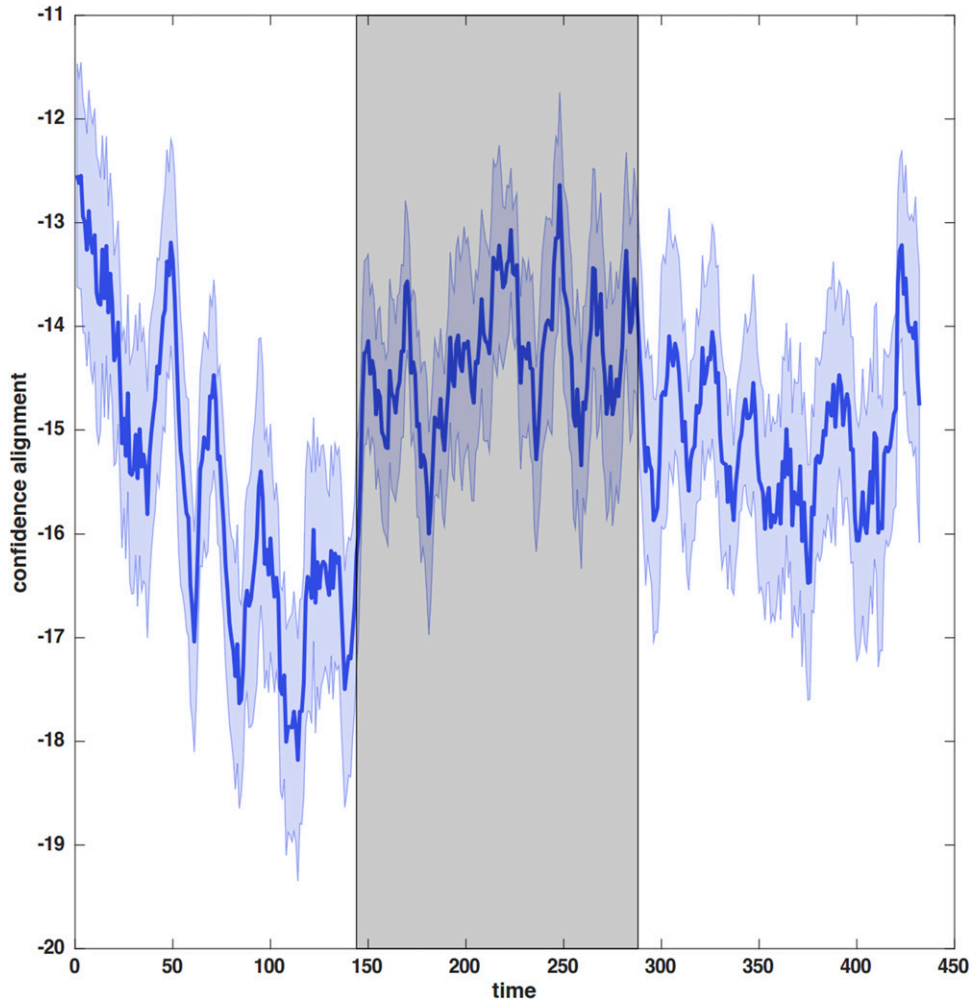


Figure 4. Sliding window average ($k = 10$) of the trial-by-trial difference between dyad's members' initial (pre-social exchange) confidence judgments over the course of the experiment. As observed, confidence difference reduced during the interactive phase 2 of the experiment (shaded area), when participants could observe each other's responses.

pre-social exchange phase. A linear regression tested whether participants tended to shift their confidence toward the most confident participant or toward the most calibrated one (Table 1). The absolute difference between phase I and II in the mean pre-advice confidence acted as dependent variable. Average confidence and confidence calibration in phase I acted as independent variables. Confidence calibration was calculated as type II A_{ROC} curves. Independent variables were scaled in the range 0–50 and can thus be interpreted as percent points from chance. Baselines refer to the variable's value at chance (50% confidence and 50% calibration), and effects refer to changes in the response variable associated with a one percent increase in the explanatory variable. Results show that smaller confidence shifts occurred when confidence was at chance and average calibration increased by 1% ($\beta = -1.0$, $SE = 0.18$, $p < .001$) or when

calibration was at chance and confidence increased by 1% ($\beta = -0.20$, $SE = 0.06$, $p = .005$). The interaction term between the two factors was significant and positive ($\beta = 0.04$, $SE = 0.007$, $p < .001$). Uncalibrated people shifted their average confidence toward their partners less the more confident they were, but calibrated people shifted their average confidence more the more confident they were. This interaction seems to suggest interindividual differences in social adjustment as a function of meta-cognitive awareness (see Kruger and Dunning, 1999). According to the model in Table 1, a hypothetical individual with average confidence and calibration 1% above chance would show an absolute confidence shift of about $8.02 - 1.02 - 0.20 + 0.04 = 6.84$ confidence points. Overall, these results suggest that confidence and calibration are necessary but not sufficient conditions for shifting confidence.

The benefits of confidence alignment

The results above establish that our interaction manipulation produced confidence alignment among dyad members, even though there was no explicit incentive for them to do so given that their decisions were registered (and rewarded) separately. We next looked at whether confidence alignment was nevertheless beneficial for group performance. Accuracy improvement was defined as the difference between pre- and post-social exchange choice accuracy during phase II, averaged across dyad members, with dyad as the random effect. A negative trend ($r(22) = -.37, p = .07$) was found between accuracy improvement and the difference between confidence means of members belonging to a same dyad during phase I

$$\Delta_C = |\bar{C}_{S1} - \bar{C}_{S2}| \quad (2)$$

providing some suggestion that greater accuracy improvements may be observed when participants' confidence distributions started off relatively aligned to each other (Figure 5).

Second, we calculated accuracy improvement after alignment on nominal dyads obtained by applying a maximum confidence slating rule (Koriat, 2012; Pescetelli et al., 2016; Bahrami et al., 2012). Nominal group accuracy was calculated by taking the accuracy of the most confident individual in the dyad on every trial in phases I (pre-interaction) and III (post-interaction), in which the participants made their decisions separately, Figure 6. In both phases, nominal group accuracy was higher than accuracy of the best-performing individual in the dyad (pre-interaction: 74.6%, $t(23) = 5.36, p < .001, d = 1.45$ and post-interaction: 76.9% $t(23) = 7.26, p < .001, d = 2.09$), reflecting the previously reported benefits of confidence-weighted agreement even without direct interaction (Koriat, 2012), with these analyses performed with dyad as the random effect. Nevertheless, compared to pre-alignment, nominal group accuracy after alignment (*i.e.* after phase II) was significantly higher ($t(23) = -2.17, p = .04, d = 0.66$), again suggesting that social interaction and alignment would have been beneficial for group performance and social coordination.

Table 1. Linear coefficients and standard errors of a multi linear model run to predict participant's absolute confidence shift from phase I to phase II of the experiment ($N = 48$). Adjusted R-squared = 0.38. F-statistic versus constant model: 10.6, $p < .001$.

| | Estimate | SE. | t-stat (d.f.) | p |
|------------------------|----------|-------|---------------|------------------|
| Intercept | 8.02 | 1.41 | 5.68 (44) | $p < .001^{***}$ |
| Calibration | -1.02 | 0.18 | -5.41 (44) | $p < .001^{***}$ |
| Confidence | -0.20 | 0.06 | -2.92 (44) | $p = .005^{**}$ |
| calibration:confidence | 0.04 | 0.007 | 5.53 (44) | $p < .001^{***}$ |

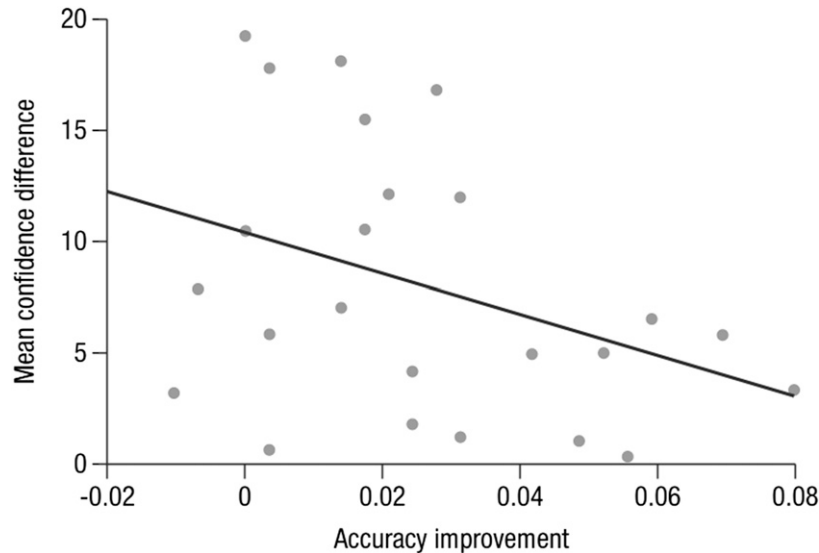


Figure 5. Negative trend existing between confidence alignment during phase I – calculated as absolute difference between participants' confidence – and accuracy improvement during phase II. The correlation suggests that greater initial mis-alignment decreases the chances of accuracy improvement at interaction.

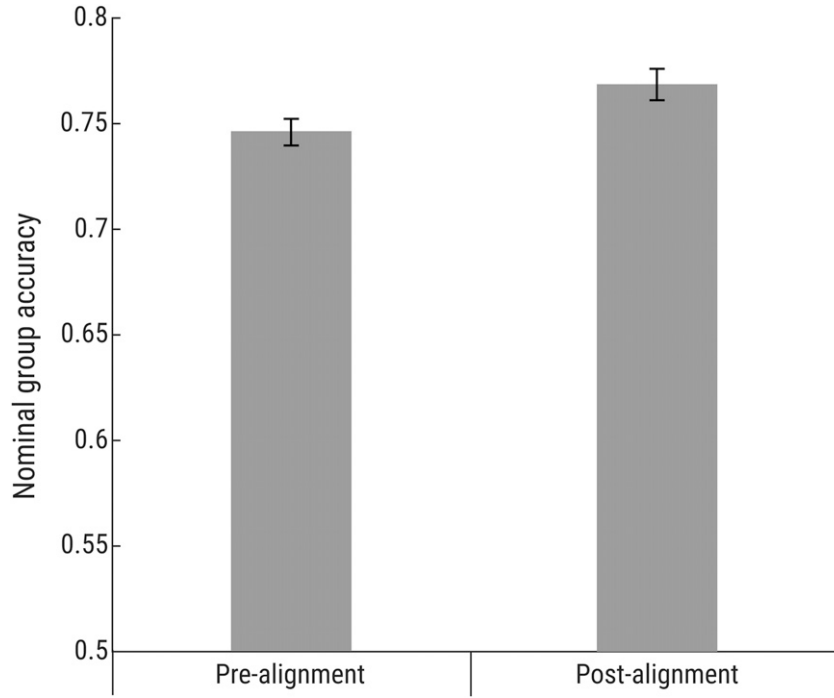


Figure 6. Accuracy improvement after alignment. Nominal group accuracy was calculated by taking the accuracy of the most confident individual in the dyad. As observed nominal group accuracy was significantly greater after alignment (phase 3) than before alignment (phase 1).

Whereas the analysis of Figure 6 estimates accuracy improvement at the group level, a final analysis estimated the accuracy benefits of confidence alignment for individual decision makers. We modelled each participant's phase II responses following social interaction as if they had not aligned confidence with their partner, using values taken from each participant's observed responses (i.e. with no further free parameters). Specifically, we estimated how each participant would have responded in phase II had they not shifted their confidence distribution based on the dyadic interaction. Thus, we applied the empirically observed confidence shift in phase II (after alignment) to distribution coordinates from phase I (pre-alignment) (Figure S2). Consider, for example, a participant who was under-confident in phase I, but then in phase II shifted to align with a more-confident partner. Now being more confident they may, in the face of disagreement with their partner on a given trial, stick with their initial decision more readily than if they hadn't increased their confidence via alignment (Figure S2). Conversely, an initially over-confident decision maker should be more likely to change their minds after having aligned with a less confident partner when, on a given trial, that partner disagrees with their own initial choice. We assessed these impacts of confidence alignment using information about participants' initial-decision confidence distributions in phases I and II, together with their

confidence shift on each trial after interaction. Using this information, we can model participants' final decisions as if had they failed to align with their partners during phase II. Specifically, for each trial, we estimate what pre-social exchange confidence \hat{C} was to be expected without alignment and then add the empirically observed confidence shift. \hat{C} was estimated as

$$\hat{C} = \mu_{pre} + \sigma_{pre} * z \quad (3)$$

where z is the z-score of pre-advice confidence observed during the interaction phase (phase II) and μ_{pre} and σ_{pre} are the mean and standard deviation of the pre-social exchange confidence distribution observed during the pre-interaction phase (phase I). To model the final accuracy of participants had they not aligned confidence, observed confidence change $\delta_C = C_{post} - C_{pre}$ was added on top of \hat{C} : $\hat{C}_{post} = \hat{C} + \delta_C$, where C_{pre} and C_{post} are the pre- and post-social exchange confidence respectively. Notice that while C_{pre} is always positive (representing the confidence in the range 0–50 in the decision made), C_{post} can assume negative values when changes of mind take place (Figure S2), reflecting that participants sometimes change their categorical decision, not just their confidence, during social interaction. Empirical distributions of confidence changes δ_C are shown in Figure S3. Adding observed confidence changes δ_C to estimated initial confidence \hat{C} resulted in an estimated post-

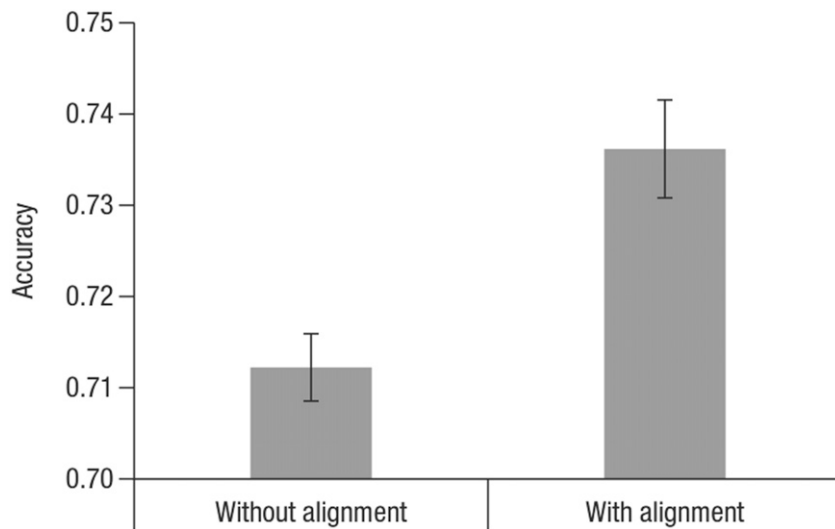


Figure 7. Accuracy that would be expected had the participants not shifted their confidence distribution toward one another (Without alignment) compared with empirically-observed performance in phase II (With alignment).

social exchange confidence according to pre-alignment coordinates, in which post-social exchange confidence values lower than zero indicate changes of mind from pre- to post-advice decisions. Thus, estimated post-social exchange accuracy \hat{C}_{or} differed from observed post-social exchange accuracy C_{or} whenever $\text{sign}(\hat{C}_{post}) \neq \text{sign}(C_{post})$, indicating changes of mind that would have not occurred without alignment (or vice versa resisting a change of mind that would have occurred without alignment) (Figure S2). Results in Figure 7 show that observed accuracy (C_{or}) was greater than modelled accuracy (\hat{C}_{or}), indicating that alignment produced greater average final accuracy than if alignment had not happened ($t(47) = 5.10, p < .001, d = 0.75$). These findings provide further support that confidence alignment produced measurable benefits on task performance and information sharing.

Discussion

The present study experimentally manipulated the interaction and social communication between two individuals performing a perceptual task in parallel. We showed that social interaction naturally elicits confidence alignment among group members, even without being prompted by the experimenter. Our contribution, compared to previous studies investigating confidence alignment (Bang et al., 2017; Fusaroli et al., 2012), is that participants in the present experiment were not performing a joint task with shared payoffs and so there was no explicit strategic incentive to align. Here, confidence alignment naturally emerged from the observation of one's partner responses. We find further that alignment emerged rapidly once

interaction began, and persisted beyond the period during which interaction was allowed.

Confidence alignment is unlikely to be due to pre-existing similarities among group members (similarity hypothesis). Although participants in our study knew each other before the experiment took place (following standard recruitment practice in similar experiments (Pescetelli et al., 2016)), a similarity hypothesis would predict that great alignment (e.g. high within-dyad correlation of confidence) should be expected from the very beginning of the experiment, namely at phase I. Instead, although we observed a weak positive correlation between dyad members' confidence even prior to interaction, we observed that our experimental manipulation produced a sudden increase in the similarity measures considered as soon as interaction among participants was introduced (Figure 4). The observed alignment was substantial, amounting to a near threefold increase in the correlation between dyad members' mean confidence and a 36% reduction in the mean difference. Participants in our experiment tended to shift their average confidence towards the more confident member of the dyad and towards the most calibrated one. These results replicate similar findings reported in Pescetelli et al. (2016) and using post-decision wagering.

Confidence alignment could be generated by spontaneous learning of the optimal strategy, by simple anchoring or both. According to the first mechanism, participants might have learnt over repeated interactions that aligning their average confidence had benefits on their personal final accuracy. Notice that according to this explanation, participants do not necessarily have to consciously learn this association, but might do so implicitly. According to the second interpretation, the simple observation of how the

other person uses the confidence scale might produce an involuntary shift in that direction. Phenomena of anchoring are well-known in psychology (Tversky and Kahneman, 1974) and are observed even when participants are well aware of the non-informativeness of the anchor (Englich et al., 2006). An interesting question for future research is therefore whether confidence alignment could be observed under even less restrictive circumstances than those studied here. In the present study, we observed alignment even when participants made individual decisions with separate payoffs, moving beyond the joint decision tasks used previously in which alignment is an effective strategic choice (Bang et al., 2017; Fusaroli et al., 2012; Pescetelli et al., 2016). Our participants were nevertheless performing the same task with the same correct answer on each trial. Building on the present design, one could ask whether alignment is observed even if dyad members knew they were making different decisions (i.e. based on different stimuli, with potentially different correct responses) or even performing different tasks (i.e. based on different kinds of evidence and requiring different responses), such that the partner's decision contained no information about the likely correct answer. Establishing boundary conditions of confidence alignment in this way would shed useful light on the computational goal and mechanism of alignment.

Relating to this question of the purpose of confidence alignment, the second and perhaps more intriguing result we reported here is that alignment improves individual accuracy and group performance. The interpretation that is offered is that confidence alignment allows the sharing of task-relevant variance (generated by variation in stimulus perception) while minimising the influence of task-irrelevant variance (generated by idiosyncratic use of the scale that reflects socio-economic status, gender and other task-irrelevant factors). Analyses done on both observed (Figure 5) and estimated values (Figures 6 and 7) lead us to conclude that confidence alignment improved group and individual accuracy. Although these analyses make assumptions (about how a group might arrive at a decision in Figure 6, and about how an individual might have changed their mind given a different starting confidence in Figure 7), they are based on descriptive statistics from individual participants' observed data rather than fitted free parameters. Interestingly, the analyses converge to similar results, namely, a 2–3% accuracy improvement after alignment. Although apparently small, this accuracy difference is notable when considering that it represents a within-subject performance improvement rather than a difference between-population averages, and that it is similar in magnitude to the improvement that accrues from dyadic performance itself, as observed in the present study and previous work both using similar perceptual tasks (Bahrami et al., 2010; Pescetelli et al., 2016) as well as more complex stimuli (Hasan et al., 2022).

An important aspect of our task, is that participants' performance was titrated to reach an average accuracy of 70% (i.e. above chance) and individual responses were independent from each other. Although, on average, people were equally accurate, confidence and accuracy showed a positive association within an individual across trials (i.e. people were *calibrated*). This is why nominal dyadic decisions – obtained by taking the decision of the most confident participant trial-wise – were higher than the highest performing individual in the dyad. Individuals could always benefit from their partner's information and thus confidence alignment favoured optimal confidence sharing. In other words, confidence alignment was the appropriate strategy to adopt, given our experimental design.

Confidence alignment might be dependent on the presence of a common ground truth between dyad members and aligned incentives. However, prior work suggests that mere social exposure might be sufficient for confidence alignment to take place. Pescetelli et al. (2016) tested a similar 2-alternative forced choice perceptual task but added a condition where the ground truth was different for the two subjects ('conflict condition') and one where there was no ground truth ('null condition'). Confidence distributions of members of the same dyad were remarkably similar (e.g. Figures S5 and S7), suggesting that the presence of a common ground truth might not be a necessary condition for confidence alignment to emerge. Yet, some unresolved questions remain. The effect of alignment in cases where individual accuracy levels are different between members (Bahrami et al., 2010) or below chance (Koriat, 2012) remains unknown. In these scenarios, confidence alignment might counter-productively produce declines of individual performance due to equality biases and over-reliance on inaccurate individuals (Mahmoodi et al., 2015). An interesting and challenging development of the current paradigm might selectively manipulate the correlation of individual judgments and/or the amount of evidence provided to participants (Pescetelli and Yeung, 2021). In these cases, confidence alignment should produce benefits when performance among individual members is comparable and declines in performance when individuals differ in their accuracies and accurate participants are more confident than inaccurate ones (Kruger and Dunning, 1999).

Finally, the fact that social interaction led participants to shift their personal confidence distributions (and shift magnitude was modulated by metacognitive awareness) suggests potential interventions to improve confidence calibration in individuals lacking metacognitive insights and clinical populations. Metacognitive deficits are known to be important aspects of several psychiatric conditions (David et al., 2012). Future avenues of investigation could pair human participants with virtual social agents showing complementary confidence distributions compared to humans. Our findings suggest that interaction with

complementary virtual tutors (e.g. over-confident participants interacting with under-confident tutors) might improve metacognitive deficit. Future studies will need to address whether confidence shifts are temporary or stable over time and whether they affect internal monitoring ability or only the way participants use the confidence scale to report their internal uncertainty.

Conclusions

Phenomena of behavioural contagion are well-known in the psychology and biology literature. Here, we showed that in an information-sharing social decision task, interaction caused confidence alignment between individual members and alignment produced measurable benefits on individual and group-level performance. The finding hints to potential applications of confidence alignment effects to improve calibration by social interaction.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the part of a doctoral program (N.P.), and it was funded by a scholarship that was kindly provided by the Clarendon fund, Christ Church College and the Department of Experimental Psychology, University of Oxford to the first author of this work.

Data accessibility statement

 This article earned Open Data, and Open Materials badges through the study materials can be accessed at

Data and code to reproduce analysis and figures is available on OSF: Pescetelli, N., & Yeung, N. (2022, February 13). Benefits of Spontaneous Confidence Alignment Between Dyad Members. Retrieved from osf.io/v7wcm

ORCID iDs

Niccolò Pescetelli  <https://orcid.org/0000-0002-8826-2202>

Nick Yeung  <https://orcid.org/0000-0003-1905-2129>

Supplemental Material

Supplemental material for this article is available online.

References

Ais J, Zylberberg A, Barttfeld P, et al. (2016) Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* 146: 377–386. DOI: [10.1016/j.cognition.2015.10.006](https://doi.org/10.1016/j.cognition.2015.10.006)

006. URL <http://linkinghub.elsevier.com/retrieve/pii/S0010027715300846>

Bahrami B, Olsen K, Bang D, et al. (2012) What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1594): 1350–1365. DOI: [10.1098/rstb.2011.0420](https://doi.org/10.1098/rstb.2011.0420). URL <http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2011.0420>

Bahrami B, Olsen K, Latham PE, et al. (2010) Optimally interacting minds. *Science (New York, N.Y.)* 329(5995): 1081–1085. DOI: [10.1126/science.1185718](https://doi.org/10.1126/science.1185718). URL <http://www.ncbi.nlm.nih.gov/pubmed/20798320>

Bang D, Aitchison L, Moran R, et al. (2017) Confidence matching in group decision-making. *Nature Human Behaviour* 1(0117): 1–7. DOI: [10.1038/s41562-017-0117](https://doi.org/10.1038/s41562-017-0117)

Barber BM and Odean T (2001) Boys will be boys: gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics* 116(1): 261–292.

Boldt A and Yeung N (2015) Shared neural markers of decision confidence and error detection. *Journal of Neuroscience* 35(8): 3478–3484. DOI: [10.1523/JNEUROSCI.0797-14.2015](https://doi.org/10.1523/JNEUROSCI.0797-14.2015). URL <http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0797-14.2015>

Bonabeau E, Dorigo M and Theraulaz G (1999) *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press.

Broihanne M, Merli M and Roger P (2014) Overconfidence, risk perception and the risk-taking behavior of finance professionals. *Finance Research Letters* 11(2): 64–73. DOI: [10.1016/j.frl.2013.11.002](https://doi.org/10.1016/j.frl.2013.11.002). URL <http://linkinghub.elsevier.com/retrieve/pii/S1544612313000627>

Buchan NR, Croson RT and Solnick S (2008) Trust and gender: an examination of behavior and beliefs in the investment game. *Journal of Economic Behavior & Organization* 68(3–4): 466–476. DOI: [10.1016/j.jebo.2007.10.006](https://doi.org/10.1016/j.jebo.2007.10.006). URL <http://linkinghub.elsevier.com/retrieve/pii/S016726810800139X>

Campbell WK, Goodie AS and Foster JD (2004) Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making* 17(4): 297–311. DOI: [10.1002/bdm.475](https://doi.org/10.1002/bdm.475). URL <http://doi.wiley.com/10.1002/bdm.475>

David AS, Bedford N, Wiffen B, et al. (2012) Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367(1594): 1379–1390. DOI: [10.1098/rstb.2012.0002](https://doi.org/10.1098/rstb.2012.0002). URL <http://www.ncbi.nlm.nih.gov/pubmed/22492754>

Dezechache G, Conty L, Chadwick M, et al. (2013) Evidence for unintentional emotional contagion beyond dyads. *Plos One* 8(6): e67371. DOI: [10.1371/journal.pone.0067371](https://doi.org/10.1371/journal.pone.0067371). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3696100&tool=pmcentrez&rendertype=abstract>

Englich B, Mussweiler T and Strack F (2006) Playing dice with criminal sentences: the influence of irrelevant anchors on experts'

- judicial decision making. *Personality and Social Psychology Bulletin* 32(2): 188–200. DOI: [10.1177/0146167205282152](https://doi.org/10.1177/0146167205282152).
- Fleming SM and Lau HC (2014) How to measure metacognition. *Frontiers in Human Neuroscience* 8: 443. DOI: [10.3389/fnhum.2014.00443](https://doi.org/10.3389/fnhum.2014.00443). URL http://www.frontiersin.org/Human_Neuroscience/10.3389/fnhum.2014.00443/abstract
- Frith CD (2007) The social brain? *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* 362(1480): 671–678. DOI: [10.1098/rstb.2006.2003](https://doi.org/10.1098/rstb.2006.2003). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1919402&tool=pmcentrez&rendertype=abstract>
- Fusaroli R, Bahrami B, Olsen K, et al. (2012) Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science Online*: 1–9. DOI: [10.1177/0956797612436816](https://doi.org/10.1177/0956797612436816). URL <http://www.ncbi.nlm.nih.gov/pubmed/22810169>
- Gallup AC, Hale JJ, Sumpter DJT, et al. (2012) Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences of the United States of America* 109(19): 7245–7250. DOI: [10.1073/pnas.1116141109](https://doi.org/10.1073/pnas.1116141109). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3358867&tool=pmcentrez&rendertype=abstract>
- Handegard NO, Boswell KM, Ioannou CC, et al. (2012) The Dynamics of coordinated group hunting and collective information transfer among schooling prey. *Current Biology* 13: 1213–1217. DOI: [10.1016/j.cub.2012.04.050](https://doi.org/10.1016/j.cub.2012.04.050).
- Hasan E, Eichbaum Q, Seegmiller AC, et al. (2022) Improving medical image decision-making by leveraging metacognitive processes and representational similarity. *Topics in Cognitive Science* 14(2): 400–413. DOI: [10.1111/TOPS.12588](https://doi.org/10.1111/TOPS.12588). URL <https://onlinelibrary.wiley.com/doi/full/10.1111/tops.12588>
- Hein AM, Rosenthal SB, Hagstrom GI, et al. (2015) The evolution of distributed sensing and collective computation in animal populations. *eLife* 4: e10955. DOI: [10.7554/eLife.10955](https://doi.org/10.7554/eLife.10955).
- Heyes C (2017) When does social learning become cultural learning? *Developmental Science* 20(2): e12350. DOI: [10.1111/desc.12350](https://doi.org/10.1111/desc.12350). URL <http://doi.wiley.com/10.1111/desc.12350>
- Huq SF, Garety PA and Hemsley DR (1988) Probabilistic judgements in deluded and non-deluded subjects. *The Quarterly Journal of Experimental Psychology Section A* 40(4): 801–812. DOI: [10.1080/14640748808402300](https://doi.org/10.1080/14640748808402300). URL <http://www.tandfonline.com/doi/abs/10.1080/14640748808402300>
- Kennedy J, Eberhart RC and Shi Y (2001) *Swarm Intelligence*. S. Francisco: Morgan Kaufmann.
- Kiani R, Corthell L and Shadlen M (2014) Choice certainty is informed by both evidence and decision time. *Neuron* 84(6): 1329–1342. DOI: [10.1016/j.neuron.2014.12.015](https://doi.org/10.1016/j.neuron.2014.12.015). URL <http://linkinghub.elsevier.com/retrieve/pii/S0896627314010964>
- Koriat A (2012) When are two heads better than one and why? *Science (New York, N.Y.)* 336(6079): 360–362. DOI: [10.1126/science.1216549](https://doi.org/10.1126/science.1216549). URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1216549>. <http://www.ncbi.nlm.nih.gov/pubmed/22517862>
- Kramer ADI, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790. DOI: [10.1073/PNAS.1320040111](https://doi.org/10.1073/PNAS.1320040111). URL <http://www.pnas.org/cgi/doi/10.1073/pnas.1320040111>
- Kruger J and Dunning D (1999) Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology* 77(6): 1121–1134.
- Levitt H (1971) Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America* 49, 467, 477(2): Suppl 2:467+. URL <http://www.ncbi.nlm.nih.gov/pubmed/5541744>
- Mahmoodi A, Bang D, Olsen K, et al. (2015) Equality bias impairs collective decision-making across cultures. *Proceedings of the National Academy of Sciences* 112: 201421692. DOI: [10.1073/pnas.1421692112](https://doi.org/10.1073/pnas.1421692112). URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1421692112>
- Mann L (1998) Cross-cultural differences in self-reported decision-making style and confidence. *International Journal of Psychology* 33(5): 325–335. DOI: [10.1080/002075998400213](https://doi.org/10.1080/002075998400213). URL <http://doi.wiley.com/10.1080/002075998400213>
- Meyniel F, Sigman M and Mainen Z (2015) Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88(1): 78–92. DOI: [10.1016/j.neuron.2015.09.039](https://doi.org/10.1016/j.neuron.2015.09.039). URL <http://linkinghub.elsevier.com/retrieve/pii/S0896627315008284>
- Migdal P, Raczaszek-Leonardi J, Denkiewicz M, et al. (2012) Information-sharing and aggregation models for interacting minds. *Journal of Mathematical Psychology* 56(6): 417–426. DOI: [10.1016/j.jmp.2013.01.002](https://doi.org/10.1016/j.jmp.2013.01.002). URL <http://linkinghub.elsevier.com/retrieve/pii/S0022249613000096>
- Milgram S, Bickman L and Berkowitz L (1969) Note on the drawing power of crowds of different size. *Perspectives in Social Psychology* 13: 79–82.
- Moiseff A and Copeland J (2010) Firefly synchrony: a behavioral strategy to minimize visual clutter. *Science* 329(5988): 181–181. DOI: [10.1126/science.1190421](https://doi.org/10.1126/science.1190421). URL <http://www.sciencemag.org/cgi/doi/10.1126/science.1190421>
- Pescetelli N, Rees G and Bahrami B (2016) The perceptual and social components of metacognition. *Journal of Experimental Psychology: General* 145(8): 949–965. DOI: [10.1037/xge0000180](https://doi.org/10.1037/xge0000180)
- Pescetelli N and Yeung N (2020) The effects of recursive communication dynamics on belief updating. *Proceedings of the Royal Society B: Biological Sciences* 287(1931): 20200025. DOI: [10.1098/rspb.2020.0025](https://doi.org/10.1098/rspb.2020.0025). URL <https://royalsocietypublishing.org/doi/10.1098/rspb.2020.0025>
- Pescetelli N and Yeung N (2021) The role of decision confidence in advice-taking and trust formation. *Journal of Experimental Psychology: General* 150(3): 507–526. DOI: [10.1037/xge0000960](https://doi.org/10.1037/xge0000960). URL <http://arxiv.org/abs/1809.10453>. <http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000960>
- Pouget A, Drugowitsch J and Kepecs A (2016) Confidence and certainty: distinct probabilistic quantities for different goals.

- Nature Neuroscience* 19(3): 366–374. DOI: [10.1038/nn.4240](https://doi.org/10.1038/nn.4240). URL <http://www.nature.com/doifinder/10.1038/nn.4240>
- Sorkin RD, Hays CJ and West R (2001) Signal-detection analysis of group decision making. *Psychological Review* 108(1): 183–203. DOI: [10.1037/0033-295X.108.1.183](https://doi.org/10.1037/0033-295X.108.1.183). URL <http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.108.1.183>
- Tomasello M (2009) *Why We Cooperate*. Cambridge, MA: MIT Press.
- Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–1131. URL <http://www.jstor.org/stable/1738360Copy>
- Valeriani D, Cinel C and Poli R (2017) Group augmentation in realistic visual-search decisions via a hybrid brain-computer interface. *Scientific Reports* 7(1): 7772. DOI: [10.1038/s41598-017-08265-7](https://doi.org/10.1038/s41598-017-08265-7)
- Wheeler L (1966) Toward a theory of behavioral contagion. *Psychological Review* 73(2): 179–192. DOI: [10.1037/h0023023](https://doi.org/10.1037/h0023023). URL <http://content.apa.org/journals/rev/73/2/179>
- Yates J, Price PC, Lee JW, et al. (1996) Good probabilistic forecasters: the ‘consumer’s’ perspective. *International Journal of Forecasting* 12(1): 41–56. DOI: [10.1016/0169-2070\(95\)00636-2](https://doi.org/10.1016/0169-2070(95)00636-2)
- Yeung N and Summerfield C (2012) Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367(1594): 1310–1321. DOI: [10.1098/rstb.2011.0416](https://doi.org/10.1098/rstb.2011.0416). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3318764&tool=pmcentrez&rendertype=abstract>
- Zajonc RB, Adelman PK, Murphy ST, et al. (1987) Convergence in the physical appearance of spouses. *Motivation and Emotion* 11(4): 335–346. DOI: [10.1007/BF00992848](https://doi.org/10.1007/BF00992848). URL <http://link.springer.com/10.1007/BF00992848>
- Zylberberg A, Fetsch CR and Shadlen MN (2016) The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife* 5: e17688(OCTOBER2016). DOI: [10.7554/eLife.17688](https://doi.org/10.7554/eLife.17688)

Author biographies

Niccolò Pescetelli is an Assistant Professor of Cyberpsychology at the New Jersey Institute of Technology. He leads the Collective Intelligence Lab.

Nick Yeung is Tutorial Fellow in Psychology at University College, Oxford and Professor of Cognitive Neuroscience in the Department of Experimental Psychology at the University of Oxford.