Social influences on metacognitive evaluations: The power of non-verbal cues

Terry Eskenazi[1*], Benoît Montalan[3*], Amélie Jacquot[2], Joëlle Proust[4], Julie Grèzes[1], Laurence Conty[2]

[1]Laboratoire de Neurosciences Cognitives U940 INSERM, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

[2] Laboratoire de Psychopathologie et Neuropsychologie EA 2027, Université Paris 8, France

[3] Laboratoire de Psychologie et Neurosciences EA 4699, Normandie Université, 76821 Mont-Saint-Aignan, France

[4] Institut Jean Nicod, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France

* The authors made equal contributions to the manuscript.

Corresponding author :

Terry Eskenazi
terry.eskenazi@ens.fr
Laboratoire de Neurosciences Cognitives U940 INSERM
Ecole Normale Supérieure
29 rue d'Ulm, 75005 Paris, France
+31(0)761577036

Running title: Social influence and metacognition

Number of words: 5336 (excl. abstract)

Abstract

Metacognitive evaluations refer to the processes by which people evaluate their own cognitive operations with respect to their current goal. Little is known about whether this process is susceptible to social influence. Here we investigate whether non-verbal social signals automatically influence metacognitive evaluations. We asked people to perform a two-alternative forced-choice task, which was followed by a face randomly gazing towards or away from the response chosen by the participant. Participants then provided a metacognitive evaluation of their response by rating their confidence in their answer. In Experiment 1, the participants were told that the gaze direction was irrelevant to the task purpose and were advised to ignore them. The results revealed an effect of implicit social information on confidence ratings even though the gaze direction was random and therefore unreliable for task purposes. In addition, non-social cues (car, instead of face) did not elicit this effect. In Experiment 2 the participants were led to believe that cue direction (face or car) reflected a previous participant's response to the very same question, in other words the social information provided by the cue was made explicit, yet still objectively unreliable for the task. The results showed a similar social influence on confidence ratings, but with an increased magnitude than in the first experiment, and observed for both cues (car and face). We additionally showed in Experiment 2, that social information further impaired metacognitive accuracy. Together our results strongly suggest an involuntary susceptibility of metacognitive evaluations to non-verbal social information, even when it is implicit (Exp 1) and unreliable (Exp 1&2).

Keywords: Metacognition, decision confidence, gaze direction, fluency, social influence

1. Introduction

As social agents, we are highly sensitive to our conspecifics, effortlessly monitoring where they look and what they see, think or do. This automatic sensitivity has adaptive value, as observing others' interactions with the world provides us with potentially vital information about the environment. In particular, following another's gaze, often described as an automatic, reflexive behaviour (Langton, Watt & Bruce, 2000) causes rapid attention shifts in the observer to the same location at which the observed agent is attending (Tipples, 2002). This informs the observer about the object of the gazing agent's attention, from which further information regarding their intentions or other mental states can be derived (Baron-Cohen, 1995; Materna, Dicke & Their, 2008). Especially after establishing eye contact gaze can serve as an ostensive cue, allowing the observer to predict that the direction of the observed agent's attention is going to be socially significant (Csibra & Gergely, 2009). Gaze following is thus a powerful mechanism through which we acquire information about the world (Becchio, Bertone & Castiello, 2008; Csibra & Gergely, 2009; Feinman Roberts, Hsieh, Sawyer & Swanson, 1992).

Recent studies have demonstrated that automatic shifts of attention in response to others' gaze behaviour can affect our evaluations about objects in the environment (Bayliss, Frischen, Fenske & Tipper, 2007; Bayliss, Paul, Cannon & Tipper, 2006; Reid & Striano, 2005) in a similar manner to the opinions of others (Campl-Meiklejohn, Bach, Roepstorff, Dolan & Frith, 2010). In two studies, Bayliss and colleagues presented participants with a face looking towards or away from a common object. Those that were looked at were later rated more likeable than those that were not (e.g., Bayliss et al., 2006; Bayliss et al., 2007; Treinen, Corneille & Luypaert, 2012). A follow up study manipulated the facial expression of the observed face. The objects gazed at with a happy face were rated as more favourable than those gazed at with a disgusted face (Bayliss et al., 2007). These findings suggest that the salience, as well as the value we associate with an object may change as a result of an observed interaction between another person and that object (for similar results, see Hayes, 2007, also cf. Becchio et al., 2008). Observed social cues thus routinely influence our appraisal of the world.

Forming associations between observed individuals and the objects with which they interact is thought to be one of the processes involved in social learning (Heyes, 2012). An interesting question is whether such social influence on our judgements, caused by a reflexive processing of a non-verbal social cue (i.e. gaze), impacts our metacognitive evaluations of those judgments.

Metacognition refers to the processes by which one's cognitive operations are monitored and controlled (Proust, 2010). As we form a decision, we concurrently monitor our mental activity in order to assess the validity of that decision. In experimental work, metacognitive evaluations are often measured by a second-order decision, usually in the form of confidence judgments of the accuracy of one's response on the first-order task (Kepecs & Mainen, 2013). Social influences on decision making have long been established by conformity studies: humans tend to change their decisions or behaviour to match those of others (e.g. Asch, 1951). Although sparse, there is some evidence to indicate that metacognition may also be susceptible to social influence. For instance, in one study, participants made metacognitive judgments (expressed their feeling of knowing) about the material they had previously learned as they received information about the recall performance of a reference group (de Carvalho Filho & Yuzawa, 2001). Participants' confidence in future performance (expressed in a predictive judgment) was modulated by how well the group majority recalled the material. Research on advice taking, similarly, has demonstrated that revising one's position on general knowledge questions after learning about another individual's response may result in an increase in decision confidence (e.g. Kaliuzhna, Chambon, Franck, Testud & van der Henst, 2012). These findings suggest that what others think when expressed verbally may, not only cause us to alter our decisions and conform to the group, but also affect how confident we are in those decisions. One question that follows from this is whether non-verbal social information would reveal similar effects. In the studies mentioned above, social feedback was presented in the form of explicit verbal information. It is well known, however, that social interactions rely, in great part, on the decoding of non-verbal signals (Mehrabian, 1971). Moreover, these signals are powerful enough to influence our evaluations of the world. Therefore, the first question we aim to investigate is whether non-verbal signals, i.e., eye gaze, influence metacognitive evaluations of our prior decisions.

The second question we address concerns the effects of social feedback on metacognitive accuracy: How reliably do subjective confidence ratings reflect the validity of decisions? One of the functions served by confidence is to inform and thus to improve the quality of our decisions. When we feel uncertain about a decision we have just made, we seek further evidence to reduce our level of uncertainty. We then revise our decisions, thus increasing our level of confidence (Yaniv & Milyavsky, 2007). When a normative pressure to 'fit in' with others is not present, this epistemic motivation—to achieve a reliable representation of the world—is thought to be the main driving force for aligning one's own decisions with that of others, i.e. conformity behaviour (Deustch & Gerard, 1955). It is well known that the more uncertain people are, the more likely they are to revise their opinions according to those of others' (Festinger, 1954; Laland, 2004). Crucially, however, sensitivity to social influence pays off only when the information received by the social source is correct. Only then does social influence serve to improve the quality of decisions and reliably increase the level of confidence in that decision (Baranski & Petrusic, 1994). Inaccurate information, on the other hand, would mislead the decisions (Koriat, 2012), yet still provide the individual with an inappropriately increased sense of confidence (Fleming, Weil, Nagy, Dolan & Rees, 2010). In summary, unreliable social feedback may jeopardize metacognitive accuracy.

The aim of this study was to investigate whether metacognitive evaluations (i.e. retrospective judgments of confidence in one's prior decision) would automatically be influenced by social information when elicited by a non-verbal social cue such as gaze direction, even when the cue was entirely irrelevant for task purposes. We further tested whether such a potential influence may further impair metacognitive accuracy. To address these questions, we developed a paradigm that included a first-order perceptual decision task (two-alternative forced-choice, 2AFC), which was followed by a subjective rating of decision confidence (Figure 1). After making their decision on the first-order task, and before providing a confidence rating, participants were presented with either a face, or a non-social cue (i.e. car), orienting towards one of two alternative responses. The cue randomly pointed either at the response chosen by the participant (congruent condition) or the other response (incongruent condition). Crucially, the participants were explicitly instructed to ignore these cues as they were not directly relevant to the task purposes. We also manipulated the difficulty of the first-order task, in order to test whether social influence on decision confidence would depend on one's initial confidence level. It is known that decision confidence, just as accuracy, tends to scale inversely with task difficulty (Oppenheimer, 2008). In the same vein, as shown by the social psychology literature, uncertainty makes people more susceptible to social influence and more likely to conform to others (Festinger, 1954, Yaniv, 2004; Kaliuzhna et al., 2012). Here we explored whether a possible involuntary social influence on decision confidence would also be modulated by the participants' level of confidence prior to the perception of the social cue.

We made four predictions. First, if the perceived gaze renders the attended response potentially significant for the observer, we would expect a modulation of confidence ratings as a function of gaze direction. Confidence in one's chosen response would thus increase as compared to a baseline condition, if the same response is perceived to be selected by another; and decrease if the participant perceives the alternative response to be selected. Second, this effect is not expected with a non-social cue, such as a car orienting towards response alternatives. Third, the automatic assessment of non-verbal social cues is expected to occur especially when an individual's level of confidence prior to the cue is low. The modulation of confidence as a result of social cueing is thus expected to scale with task difficulty, resulting in a greater change in confidence levels for difficult trials as compared to easier trials (Festinger, 1954). Finally, given that the social feedback provided was completely random, and hence unreliable, the social influence should impair how well confidence ratings reflect the accuracy of decisions (i.e. metacognitive accuracy).

2. Experiment 1

2.1 Method

Participants. Seventy-nine volunteers were randomly assigned to one of the two *Cue-type* groups (Face vs. Car). Data from seven participants were excluded from analysis because their mean accuracy was more than two standard deviations above or below the group mean. Thus, analysis included data from 36 participants in the Face group (13 males, mean age = 25.9 ± 0.9) and 36 in the Car group (11 males, mean age = 26.8 ± 1.5). All participants were right-handed with normal or corrected-to-normal vision and naive to the experimental aims. Participants gave their written informed consent and received €15 for participation.

Stimuli. The first-order 2AFC task was a number estimation task where participants judged whether target displays contained more or fewer dots than a reference display. The displays consisted of arrays of white dots (10 pixels in diameter) randomly distributed on a black disc (320 pixels in diameter), with at least 10 pixels separating one dot from each other. For target displays, the number of dots varied from 32 to 68 by increments of 4, while the number of dots was fixed at 50 for the reference display. Task difficulty was manipulated by varying the difference of dot numbers separating the target from the reference displays. This difference ranged from ± 2 to ± 18 dots in five increments yielding five levels of task difficulty. Forty different target displays were randomly generated per each level of difference, as well as 10 different reference displays.

The stimuli used for the directional cues consisted of 10 pictures of faces (5 males, 5 females) with neutral expression and 10 pictures of cars, both in full frontal and in ¾ views, in both directions. The faces were selected from the Radboud Faces Database (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & van Knippenberg, 2010), while the car pictures were chosen from a web-based collection (http://www.cars.com). All stimuli were converted to a 256 grey-level format, rescaled proportionally to a size of 640 x 640 pixels, and matched in luminance.

Procedure. Participants were tested individually in a shielded room. They were seated approximately 90cm away from a 17in LCD monitor. The stimuli were presented by E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA). Each trial began with a fixation cross of 400ms duration, followed by a brief target display of 100ms duration. Following that, the words “minus” and “plus” appeared respectively on the left and on the right half of the screen and remained until a response was received. The task was to decide whether the target display contained more or fewer dots than the reference display that was presented for 3000ms. After responding with a mouse-click, participants were asked to indicate their level of confidence in their response using a vertical scale ranging from 0 (not confident at all) to 100 (very confident) (Figure 1).

The experiment consisted of 10 blocks of 40 trials each, with the reference display presented once every 20 trials. In each block, half of the trials constituted the neutral condition in which no cues were presented between the 2AFC task and the confidence rating. In these neutral trials, a fixation cross followed the response and remained on screen for 1500ms. In the other half of the trials (congruent and incongruent), a fixation cross remained between the two response alternatives for 200ms. This was followed by a picture of a face or car, in full-frontal view for 900ms. This picture was then replaced with the same stimuli but displayed in a ¾ view (left or right) for 400ms. The succession of the two pictures (frontal and then angled) created an apparent motion where the face or the car appeared to turn towards one of the two response alternatives. For the congruent trials (25%), the stimulus turned toward the response just given by the participant, irrespective of its correctness. For the incongruent trials (25%), they turned away from the participant’s response. Incongruent and congruent trials were presented in random order, but occurred equally for each level of difficulty. In both cue groups (face and car) participants were instructed that these directional stimuli were completely uninformative with respect to the correct response to the dot question and should therefore be ignored. The experiment began with a training session of 10 trials.

Please insert Figure 1 about here

2.2. Analysis and results

Number estimation task. The accuracy and reaction times (RTs) were analysed by means of a mixed design analysis of variance (ANOVA), in a 5 x 2 factorial design with the within-subjects factor *Difficulty* (levels 1, 2, 3, 4 and 5) and the between-subjects factor *Cue-type* (face vs. car). As predicted, the results showed significant main effects of *Difficulty* on accuracy ($F(4, 272) = 1074.15$, $\varepsilon = .70$, $p < .0001$), and on RTs, ($F(4, 272) = 96.42$, $\varepsilon = .39$, $p < .0001$; see Table 1). According to the planned comparisons with Bonferroni correction, accuracy decreased (all *p*s < .05), whereas RTs increased, with task difficulty (all *p*s < .05; except between levels 1 and 2 for RTs, $p > .10$) (Table 1).

Confidence ratings. A 5 x 3 x 2 mixed design ANOVA was carried out to analyse confidence data, with within-subjects factors Difficulty and *Cue-direction* (congruent, incongruent or neutral) and with between-subjects factor *Cue-type* (face vs. car). The results yielded a significant main effect of *Difficulty* ($F(4, 272) = 251.090$, $\varepsilon =.36$, $p < .0001$). Overall confidence ratings decreased with increasing levels of task difficulty (planned comparisons with Bonferroni correction, all *p*s < .05). Importantly, a significant two-way interaction between *Cue-direction* and *Cue-type* was found ($F(2, 136) = 3.750$, $\varepsilon = .96$, $p < .05$) (Figure 2). Confidence levels did not change with *Cue-direction* in the Car group. In the Face group only confidence ratings increased when the cue was congruent when compared to incongruent and neutral trials.

Please insert Figure 2 about here

Metacognitive accuracy. We also computed participants' metacognitive accuracy for each condition. Metacognitive accuracy, in this context, refers to how well one's decision confidence reflects objective task performance. It is commonly quantified as the relation between the accuracy performance in the first-order task and the confidence rating using the Type II Receiver Operating Characteristic (ROC) curve (Fleming et al., 2010). The Type II ROC curve reflects one's ability to discriminate between one's correct and incorrect decisions on the first-order event. It is a measure of the probability of being correct on the first-order task (i.e. number estimation) for a given level of confidence. The area under the ROC curve (A_{roc}) indexes subject's metacognitive accuracy. In order to test whether social cues influence one's metacognitive accuracy, as it does one's subjective confidence, we calculated A_{roc}

values per subject per each *Cue-direction* condition (neutral, congruent, incongruent) (Galvin, Podd, Drga & Whitmore, 2003). We performed a 3 x 2 ANOVA between the factors *Cue-orientation* and *Cue-type*. The factor *Difficulty* was not included as the number of observations per condition would not be sufficient for a reliable ROC analysis. The results revealed no significant main or interaction effects (Figure 3). The orientation of the cue did not modulate metacognitive accuracy in the Face ($p > .1$) or in the Car group ($p > .1$)

Please insert Figures 3 and 4 about here

2.3. Discussion

In this experiment, we investigated whether simply observing another person gazing at one of the response alternatives in a first-order task would result in an automatic modulation of participants' confidence judgments and metacognitive accuracy. The results provided partial support for our hypotheses.

Confirming our predictions, confidence was affected by the direction of the face, and not by that of the car. That is, effects on confidence were driven by the condition in which the cue was social by nature. Thus, participants automatically adjusted their confidence ratings as a function of the information provided by gaze direction, raising their confidence when the face oriented toward the response chosen by the participant as compared to when the face oriented toward the alternative response or when no social cue was present. In contrast, this effect was not elicited by the orientation of cars. We noted that the effect was mainly driven by congruent trials. This suggests that an implicit agreement by the observed gazer resulted in a larger change in confidence than an implicit disagreement. The observed asymmetry in confidence ratings between congruent and incongruent trials is likely to be due to a confirmation bias, and is in line with the proposal that eye gaze can act more as a positive (rather than a negative) reinforcement (Bayliss et al., 2006), activating the brain's reward circuitry (Kampe, Frith & Frith, 2001). This particular sensitivity to an implicit social agreement thus converges with the view that confidence judgments partly reflect an automatic internal rewarding process (Daniel & Pollmann, 2012).

As expected, confidence also scaled with task difficulty. Confidence decreased, whereas RTs and errors increased, with increasing levels of task difficulty. This confirms that metacognitive evaluations, as indexed by retrospective confidence ratings, were sensitive to task fluency, i.e. the ease with which the information was processed (e.g., Alter & Oppenheimer, 2009; Oppenheimer, 2008). People are indeed more confident in their performance when a task is experienced as fluent rather than disfluent (e.g., Kelley & Lindsay, 1993; Koriat, 1993; Simmons & Nelson, 2006). Surprisingly, however, we did not find a modulation of social influence as a function of task difficulty. The magnitude of the social influence on confidence levels did not change depending on how confident the participants were prior to social feedback. This indicates a lack of trade-off between two types of information (internal fluency and external social). Instead, we find that both types of information contribute equally to the formation of confidence judgments.

Our results suggest a mechanism that grants a reflex-like sensitivity to social feedback during metacognitive evaluations. Note that the information provided by the cue in this experiment was entirely random and independent of the first-order task. The cue oriented towards each of the response alternatives 25% of the time, irrespective of the correct response to the question and of the response chosen by the participant. Further, the cue was presented as irrelevant to the task via explicit instructions. Therefore, our results confirm that gaze direction is automatically processed as a meaningful cue, altering the relative salience of the stimuli for the participants. The (mis)information then gets incorporated into their metacognitive evaluations, consequently modulating confidence ratings. The exact ways in which such a change in confidence would lead the participants to change their response to the first-order task, were they given an opportunity to do so, is an intriguing question, which our paradigm was not designed to address.

The results of Experiment 1 thus indicate that confidence ratings are susceptible to implicit social feedback signalled by non-verbal cues. However, this susceptibility does not seem to be powerful enough to alter one's metacognitive accuracy. Indeed, our findings further show that participants were adept at monitoring their response on the first-order task on neutral trials as well as on trials where they received social feedback (A_{roc} ps<.001 for all conditions, Fig. 3).

We thus conducted a second experiment in which the meaning of the cues was contextually enhanced. Here, the participants were instructed that the direction of the cues (face and car) reflected a previous participant's responses to the first-order task. In contrast to the previous experiment, the information provided by both cues was socially meaningful and provided explicit information. The cues, however, remained objectively unreliable for the task; the direction of the cue was randomly distributed independently of the participants' response. In addition, and importantly, the participants were not instructed to consider the feedback while monitoring their confidence. We thus still explored an implicit effect of social cues on metacognitive evaluations. We therefore expected both cues (face and car) to trigger a similar pattern of effects as revealed in the Face group of Experiment 1. Secondly, if social influence on metacognitive evaluations (decision confidence) is indeed driven by an epistemic motivation, i.e., an adaptive urge to achieve an accurate model of the world, then metacognitive accuracy should be altered by the explicit social feedback provided by both types of cues as they are both socially meaningful.

3. Experiment 2

3.1. Method

Participants. Sixty-six volunteers were randomly assigned to one of the two groups. Data from five participants were excluded from analysis because mean accuracy was more than two standard deviations above or below the group mean. Thus, data from 33 participants in the Face group (10 males; mean age = 26.5 ± 1.0) and 28 in the Car group (12 males; mean age = 25.7 ± 1.3) were analyzed. Participants presented similar characteristics as in Experiment 1, gave their written informed consent, and received payment for their participation.

Stimuli and procedure. The stimuli and the procedure in Experiment 2 were similar to those used in Experiment 1 except for instructions. Here, we informed participants that the orientation of the cues (face or car) reflected the response given by a previous participant to that particular trial. Moreover, in order to reinforce this belief manipulation, participants in the Face group were photographed before the training session and were led to believe that their photographs would be used in subsequent experimental sessions to represent the responses they gave. Participants in the Car group selected a car image they liked for the same purpose.

3.2. Results

Number estimation task. The ANOVA performed on the accuracy and RT data revealed a main effect of *Difficulty* on both accuracy ($F(4, 236) = 1046.08$, $\varepsilon = .64$, $p < .001$) and RTs ($F(4, 236) = 91.41$, $\varepsilon = .47$, $p < .001$). Planned comparisons with Bonferroni correction showed that accuracy decreased, all *p*s < .05, and RTs increased with task difficulty, all *p*s < .05 (except between difficulty levels 4 and 5) (Table 2).

Please insert Table 2 about here.

Confidence ratings. The 5 x 3 x 2 mixed design ANOVA conducted on subjective confidence revealed, as in Experiment 1, a main effect of *Difficulty* ($F(4, 236) = 229.17$, $\varepsilon = .35$, $p < .001$). Planned comparisons with Bonferroni correction showed that confidence decreased with task difficulty (all *p*s < .05). As in Experiment 1, confidence scaled with task difficulty. The significant main effect of *Cue-direction* was also found ($F(2, 118) = 13.85$, $\varepsilon = .65$, $p < .0001$). Confidence levels were higher on congruent compared to incongruent ($F(1, 59) = 15.38$, $p < .001$) and neutral trials, ($F(1, 59) = 23.14$, $p < .001$). In addition, confidence levels were lower in incongruent compared to neutral trials ($F(1, 59) = 3.73$, $p = .058$; Figure 5). Finally, a significant two-way interaction of *Cue-direction* and *Difficulty* was found ($F(8, 472) = 3.07$, $\varepsilon = .62$, $p < .05$). For both groups of *Cue-type* (car and face), confidence was greater in congruent than neutral trials only with low levels of task difficulty (levels 1, 2 and 3, $p < .005$ for all levels). In contrast, it was lower for incongruent as compared to baseline trials for more difficult trials only (levels 4 and 5, all $p < .005$; Figure 5). Yet, the difference in confidence ratings between congruent and incongruent trials remained significant for each level of difficulty (ps < .05 for all levels) with similar effect sizes (range= 4.42-6.15, mean = 5.08 ± .65).

Please insert Figures 4 and 5 here

Metacognitive Accuracy. The ANOVA between factors *Cue-type* and *Cue-direction* on A_{roc} values revealed a main effect of *Cue-direction* ($F(2,128) = 6.734$, $p < .01$) (Figure 6). In both groups, compared to congruent and incongruent trials, metacognitive accuracy was highest with neutral trials. The main effect of *Cue-type* and the two-way interaction were not found significant (*ps* > 1).

Please insert Figure 6 here

3.3. Discussion

In this second experiment, we investigated whether observing a cue oriented toward one of the responses to a first-order task and believing that this orientation represented a previous participant's response, resulted in an automatic modulation of confidence ratings and of metacognitive accuracy. The results confirmed our hypotheses.

Regarding subjective confidence, our findings revealed a pattern similar to the one found in the Face group of Experiment 1. Participants adjusted their confidence ratings as a function of the information provided by gaze direction; raising their confidence when the face oriented toward the response chosen by the participant, lowering it when it oriented toward the non-chosen response. However, relative to Experiment 1, the effect of social feedback was larger, and marked for both incongruent and congruent trials. The fact that both experiments yielded a similar pattern of results suggests that participants were unable to fully ignore the social cues (i.e. face) in Experiment 1 and were involuntarily influenced by the social information the cues implicitly conveyed when forming their metacognitive evaluations. Participants are likely to have attributed attitudes to the gazer concerning the response toward which he or she gazed (Baron-Cohen, 1995). Hence, that response acquired new properties in the eye of the observer (Becchio et al., 2008). This converges with previous results suggesting that the value of the response alternatives is altered by the observation of gaze behaviour (Bayliss et al., 2006; cf. Becchio et al., 2008). However, in Experiment 2, both the non-social (i.e. car) and the social cue (i.e. face) orientation represented the decisions of previous participants. Car cues were as socially meaningful as face cues, and therefore yielded the same effect on confidence as did face cues. This rules out the possibility that the modulation of confidence levels observed for faces in Experiment 1 and 2 were merely due to gaze cueing effects. This result instead demonstrates that, once a cue is attributed a social meaning, it becomes relevant to metacognitive evaluations. It should be remembered that the participants in Experiment 2 were not specifically instructed to take this information into account when monitoring their confidence.

In contrast to Experiment 1, the results of Experiment 2 revealed a significant interaction between task difficulty and the direction of the cue. Congruent cues increased confidence mainly for difficult trials, whereas incongruent cues decreased confidence as the task became easier (Figure 8). However, a social influence on metacognitive evaluations was found whatever the participants' confidence level was prior to social feedback. Such a pattern of the interaction thus might reflect a mere ceiling/floor relation. The fact that we did not observe a decrease in confidence in the responses to incongruent cues in difficult trials suggests that participants were already at their most uncertain. They could not be any less confident than they already were after receiving the cue, hence the floor effect. In the same vein, participants were most confident in easier trials; therefore congruent cues could not increase participants' confidence, resulting in a ceiling effect.

Finally, as predicted, social feedback also affected participants' metacognitive accuracy (Figure 6). Participants were best at monitoring their performance when they received no feedback (neutral trials); that is, reported levels of confidence were a better marker of the accuracy of their given responses. However, as soon as they received explicit social information about previous participants' decisions, their metacognitive accuracy declined. Due to the particulars of our design, we cannot conclude that this finding would transfer to all situations. The reason we found social influence to impair metacognitive accuracy is that the feedback provided by the cue was purely random and unreliable. Therefore, any effect of social influence on metacognitive accuracy would have to be inevitably misleading. However, this finding has an important implication; it indicates that this automatic appraisal of social cues and our susceptibility to social influence can impact how well we evaluate our decisions, even at the expense of being wrong about our decisions.

4. General Discussion

Non-verbal social cues, such as gaze and emotional expressions, play a large role in our social interactions. We process them automatically, and they thus influence ongoing and subsequent processes. Here, we showed in two experiments that non-verbal social cues can involuntarily impact our metacognitive evaluations of the decisions we make (i.e. confidence). Furthermore, non-verbal social cues, when conveying explicit (mis)information, can sway the accuracy of those evaluations (i.e. metacognitive accuracy).

Our results first reveal that metacognitive evaluations are sensitive to both task difficulty and non-verbal social cues. This is in line with the view that metacognitive judgments rely on a variety of online cues (Koriat & Levy-Sadot, 2000). Fluency is well known to underpin reliable judgments about performance accuracy (e.g., Alter & Oppenheimer, 2009; Oppenheimer, 2008). However, to our knowledge, this is the first study to demonstrate that metacognitive evaluations can involuntarily rely on non-verbal social cues independent of their relevance, or of task difficulty, and importantly, in the absence of a normative pressure.

Our data do not allow us to draw any definitive conclusions regarding the nature of the underlying mechanism. We could conjecture that through non-verbal social cues the derived perceptual or mental states of an observed agent can change the value of the objects with which the agent interacts (Becchio et al., 2008). Here, we show that a non-agent (i.e. car), when attributed social meaning, can instigate a similar process. In both cases, participants automatically aligned their metacognitive evaluations with the non-verbal social feedback, just as was reported by previous studies that manipulated verbal information (e.g. Kaliuzhna et al., 2012; Yaniv et al., 2007). Our results reveal a tendency to automatically refer to other's decisions, even when they are non-verbally expressed, and objectively unreliable. Such a social influence on metacognitive evaluations, as we show, may compromise metacognitive accuracy. Although exploiting social cues as good diagnostic tools when assessing the quality of our beliefs and decisions serves as an adaptive strategy, it does come with pitfalls. Relying on others when evaluating ourselves can influence how we assess those decisions, at the expense of being misled. This is in line with findings reported by collective decision-making studies showing that, even though incorporating wrong information can lead group decisions astray, individuals can nonetheless become increasingly confident with those decisions (Koriat, 2012; Mahmoodi, Bang, Ahmadadi & Bahrami, 2013).

As highly social animals, humans are attuned to conspecifics with a strong attentional, perceptual and motivational bias towards information transmitted by them (Heyes, 2012). This bias can serve social learning, which is an adaptive and cheap way of acquiring novel information (Rendell, Fogarty, Hoppitt, Morgan & Webster, et al., 2011). However, there are situations where privileging information signalled by others can prove to be more costly than individual learning, misleading the behaviour of the learner and thus promoting the transmission of erroneous information (Rendell et al., 2011; Rendell, Boyd, Cownden, Enquist & Eriksson et al., 2010; Enquist & Eriksson, 2007; Giraldeau, Valone & Templeton, 2002). Our data suggest that our bias towards social information, when signalled through non-verbal cues, can directly and automatically mislead metacognitive evaluations. We suspect that the misguided modulation in confidence levels we found, and the consequential reduced metacognitive accuracy, reflects the revision of participant's initial decisions, i.e. behavioural change. Even though our paradigm was not designed to test this hypothesis, such a potential coupling between a social influence on decision confidence and behaviour change may have a critical role in (social) learning. In sum, our work emphasizes the importance of implicit non-verbal information in self-evaluations. The ways in which such a social influence on metacognition may impact the quality of learning remains an open question.

Acknowledgements

The research was supported by a European Research Council (ERC) advanced Grant, “Dividnorm” (Project 269616) and by two institutional grants: ANR-11-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL. We thank Asher Koriat for fruitful discussions that helped us develop this study and Valerian Chambon for his input in this manuscript.

Figures and Tables

Figure 1. Schematic illustration of an example trial. The first row represents a neutral trial (50%), the second and the third congruent or incongruent trials (25% each).

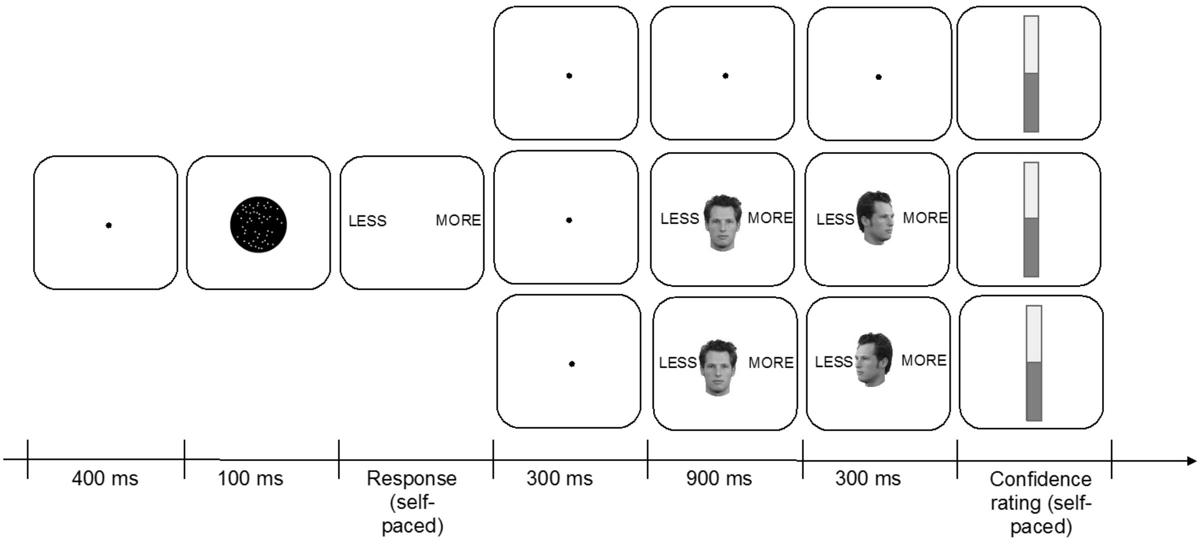

Figure 2. Subjective confidence levels in Experiment 1 as a function of the *Cue-direction* (incongruent - baseline vs. congruent-baseline) and *Cue-type* (face vs. cars). The data points of the graph are confidence values of congruent and incongruent confidence level after neutral (no cue) confidence levels are subtracted.

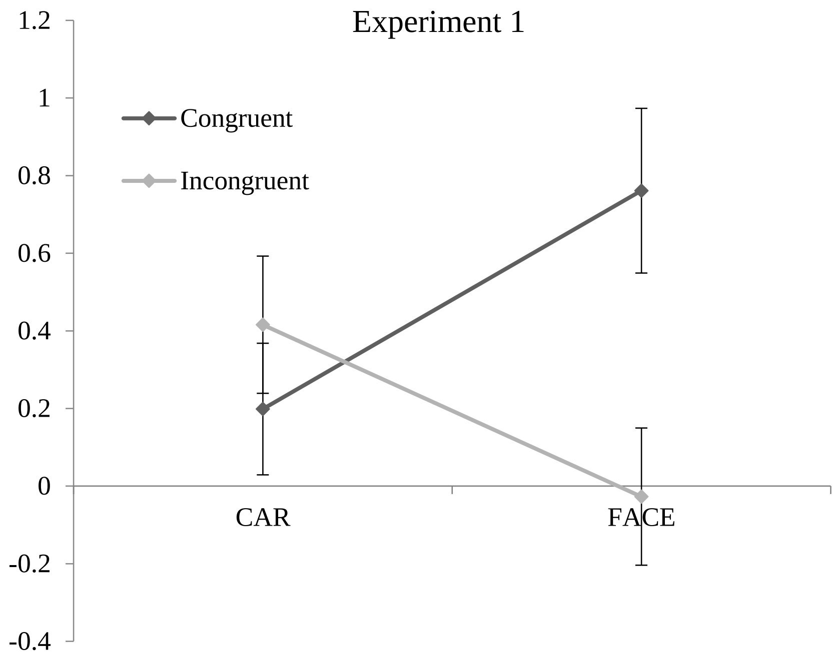

Figure 3. Mean area under the curve (A_{roc}) for non-social and social cue groups in Experiment 1.

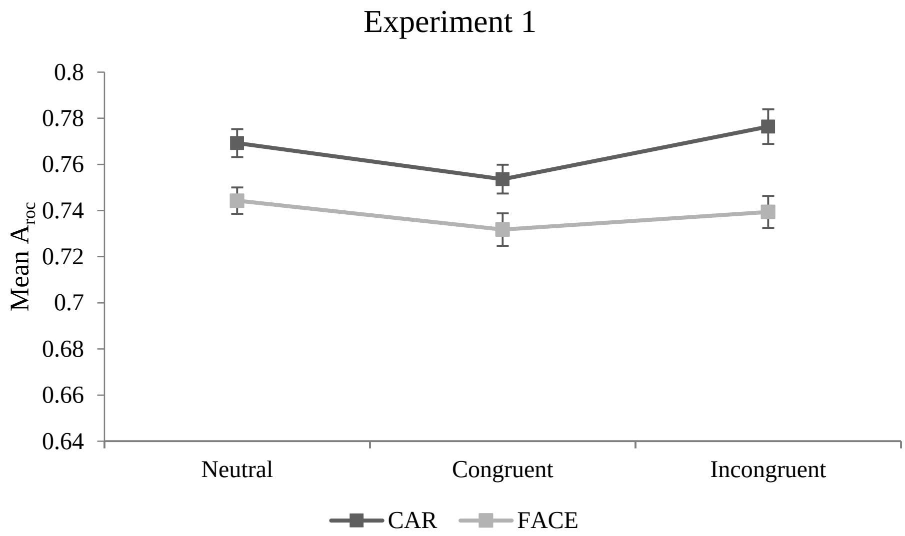

Figure 4. Subjective confidence levels in Experiment 2 plotted as a function of the *Cue-direction* (incongruent vs. baseline vs. congruent) and *Cue-type* (face vs. cars). The data points of the graph are confidence values of congruent and incongruent confidence level after neutral (no cue) confidence levels are subtracted.

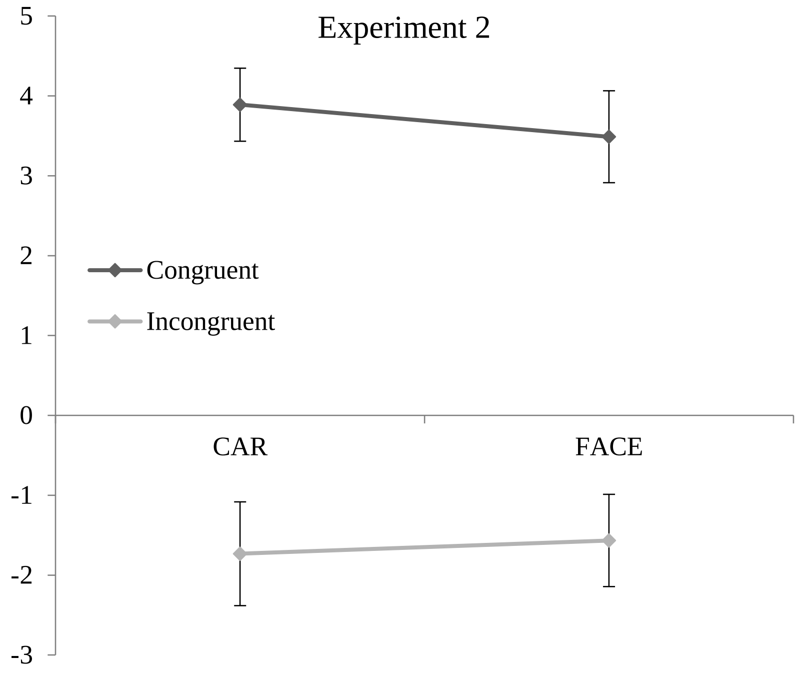

Figure 5. Change in confidence levels (congruent/incongruent-neutral) in Experiment 2 plotted as a function of *Cue-direction*, *Cue-type* and *Difficulty*.

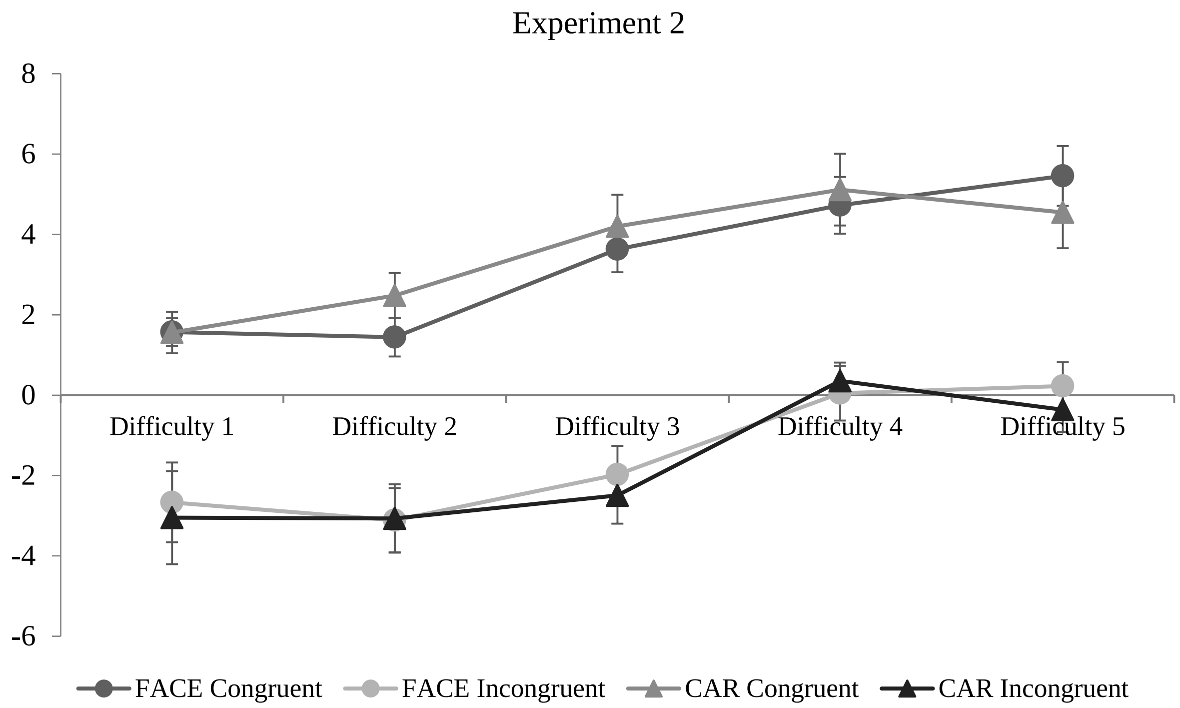

Figure 6. Mean area under the curve (A_{roc}) for non-social and social cue groups in Experiment 2.

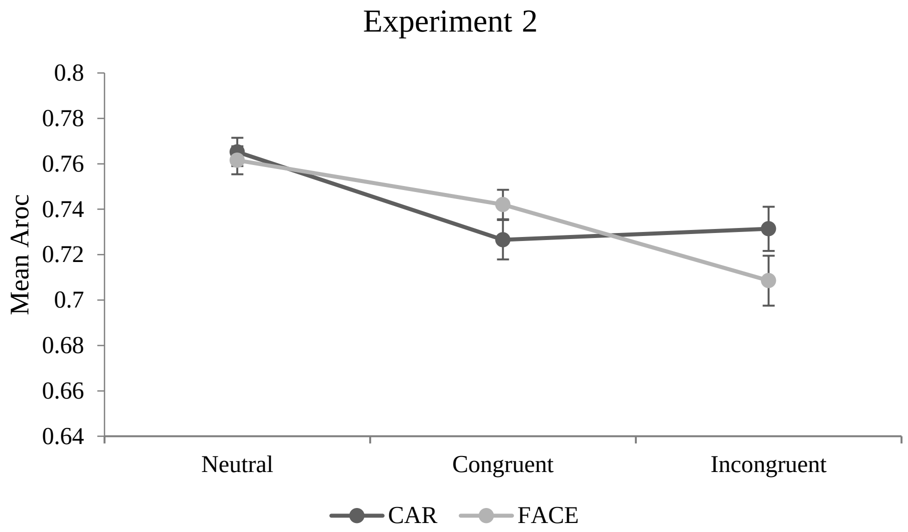

Table 1. Means (%) and standard errors for accuracy and reaction time data for Experiment 1.

	Social cue (face)				Non-social cue (car)			
	%		RTs		%		RTs	
Difficulty (distance)	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Level 5(±2)	59,74	0,85	1051,79	83,97	60,07	0,83	1022,18	81,61
Level 4(±6)	78,01	1,16	931,83	76,46	78,37	1,13	884,86	74,30
Level 3(±10)	89,78	1,04	693,58	48,17	89,79	1,01	711,89	46,81
Level 2(±14)	95,77	0,72	560,40	43,80	95,31	0,70	602,07	42,57
Level 1(±18)	97,87	0,44	513,51	34,66	97,74	0,43	513,94	33,68

Table 1. Means (%) and standard errors for accuracy and reaction time data for Experiment 2.

	Social cue (face)				Non-social cue (car)			
	%		RTs		%		RTs	
Difficulty (distance)	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Level 5(±2)	60,31	1,05	956,90	77,25	59,57	1,10	1008,08	81,15
Level 4(±6)	80,34	0,96	854,39	79,26	79,31	1,00	886,45	83,26
Level 3(±10)	91,28	0,83	639,62	43,83	90,00	0,87	715,87	46,04
Level 2(±14)	96,35	0,61	534,91	34,05	95,78	0,64	589,74	35,77
Level 1(±18)	98,13	0,36	507,21	34,06	98,33	0,38	484,67	35,78

REFERENCES

Alter, A. & Oppenheimer, D. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review, 13(3),* 219-235.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh, PA: Carnegie Press.

Baranski, JV, Petrusic, WM (1994). The calibration and resolution of confidence in perceptual judgments. *Perception Psychophysics, 55,* 412–428

Bayliss, A., Paul, M., Cannon, P. & Tipper, S. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review, 13(6),* 1061-1066.

Bayliss, A., Frischen, A., Fenske, M. & Tipper, S. (2007). Affective evaluations of objects are influenced by observed gaze direction and emotional expression. *Cognition, 104(3),* 644-653.

Baron-Cohen, S. (1995). Mindblindness: an Essay on Autism and Theory of Mind, MIT Press: Bradford Books.

Becchio, C., Bertone, C. & Castiello, U. (2008). How the gaze of others influences object processing. *Trends in Cognitive Sciences, 12(7),* 254-258.

Csibra, G. & Gergely, G. (2009). Natural Pedagogy. *Trends in Cognitive Sciences, 13(4),* 148-153.

Daniel, R. & Pollman, S. (2012). Striatal activations signal prediction errors on confidence in the absence of external feedback. *Neuroimage, 59(4),* 3457-3467.

De Carvalho Filho, M.K. & Yuzawa, M. (2001). Effect of social influences and general metacognitive knowledge on metamemory judgments. *Contemporary Educational Psychology, 26,* 571-587.

Deustch, M. & Gerard, H. (1955). A Study of Normative and Informational Social Influences upon Individual Judgment. *The Journal of Abnormal and Social Psychology, 51(3),* 629-636.

Enquist, M. & Eriksson, K. (2007). Critical Social Learning: A Solution to Rogers' Paradox of Nonadaptive Culture. *American Anthropologist, 109(4)*, 727-734.

Feinman, S., Roberts, D., Hsieh, K.F., Sawyer, D. & Swanson, K. (1992). A critical review of social referencing in infancy. In S, Feinman (Ed.), *Social Referencing and the Social Construction of Reality in Infancy*. New York: Plenum Press.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations, 7(2),* 117-140.

Fleming, S., Weil, R. & Nagy, Z, Dolan, R. & Rees, G. (2010). Relating Introspective Accuracy to Individual Differences in Brain Structure. *Science, 329,* 1541-1543.

Galvin, S.J., Podd, J.V., Drga, V. & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review, 10(4),* 843-876.

Giraldeau, L-A., Valone, T.J. & Templeton, J.J. (2002). Potential disadvantages of using socially acquired information. *Philosophical Transactions of The Royal Society B, 357,* 1559-1566.

Heyes, C. (2012). What's Social About Social Learning? *Journal of Comparative Psychology, 126(2),* 193-202.

Kaliuzhna, M., Chambon, V., Franck, N., Testud, B. & Van der Henst, J-B. (2012). Belief Revision and Delusions: How do patients with schizophrenia take advice? *PLoS one, 7(4),* e34771.

Kampe, K., Frith, C. & Frith, U. (2003). "Hey John": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing", regardless of modality. *The Journal of Neuroscience, 23(12),* 5258-5263.

Kelley, C. M. & Lindsay, D. S. (1993). *Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions.* Journal of Memory and Language, 32, 1–24.

Kepecs, A. & Mainen, Z. (2013). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B, 367,* 1322-1337.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling-of-knowing. *Psychological Review, 100,* 609-639.

Koriat, A. & Levy-Sadot, R. (2000). Conscious and unconscious metacognition: A rejoinder. *Consciousness and Cognition, 9,* 193-202.

Koriat, A. (2012). When are two heads better than one and why? *Science, 336* (6079), 360-362.

Laland, K.N., (2004). Social learning strategies. *Learning Behaviour, 32,* 4-14.

Langton, S., Watt, R. & Bruce, I. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4(2),* 50-59.

Mahmoodi, A., Bang, D., Ahmadabadi, M.N. & Bahrami, B. (2013). Learning to make collective decisions: the impact of confidence escalation. *PLoS One, 8(12),* e82295.

Materna, S., Dicke, P. & Their, P. (2008). Dissociable roles of superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *Journal of Neuroscience, 20(1),* 108-109.

Mehrabian, A. (1971). Nonverbal betrayal of feeling. *Journal of Experimental Research in Psychology, 5,* 64–73.

Oppenheimer, D. (2008). The secret life of fluency. *Trends in Cognitive Sciences, 12(6),* 237-241.

Proust, J. (2010). Metacognition. *Philosophy Compass, 5(11),* 989-998.

Reid, V.M. & Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *European Journal of Neuroscience, 21*, 11763-1766.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M.W., Fogarty, L., Ghirlande, S., Lillicrap, T. & Laland, K.N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science, 328(5975)*, 208-213.

Rendell, L., Fogarty, L., Hoppitt, W.J.E., Morgan, T.J.H., Webster, M.M. & Laland, K.N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences, 15(2),* 68-76.

Simmons, J. & Nelson, L. (2006). Intuitive confidence: choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General, 135(3),* 409-428.

Tipples, J. (2002). Eye gaze is not unique: automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review, 9(2),* 314-318.

Treinen, E., Corneille, O. & Luypaert, G. (2012). L-eye to me: the combined role of Need for Cognition and facial trustworthiness in mimetic desires. *Cognition, 122,* 247-251.

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.

Yaniv, I. (2004). Receiving other people’s advice: Influence and benefit. *Organizational Behavior and Human Decision Processes, 93(1),* 1-13.