Consciousness and Cognition 27 (2014) 246–253

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Variance misperception explains illusions of confidence in simple perceptual decisions

Ariel Zylberberg [a,b,c,d,*], Pieter R. Roelfsema [b,e,f], Mariano Sigman [a,g]

[a] Laboratory of Integrative Neuroscience, Physics Department, FCEyN UBA and IFIBA, Conicet, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina
[b] Department of Vision and Cognition, Netherlands Institute for Neuroscience, an Institute of the Royal Netherlands Academy of Arts and Sciences, Meibergdreef 47, 1105 BA Amsterdam, The Netherlands
[c] Institute of Biomedical Engineering, Faculty of Engineering, Buenos Aires University, 1063 Buenos Aires, Argentina
[d] Laboratory of Applied Artificial Intelligence, Computer Science Department, FCEyN UBA, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina
[e] Department of Integrative Neurophysiology, VU University Amsterdam, The Netherlands
[f] Psychiatry Department, Academic Medical Center, Amsterdam, The Netherlands
[g] Universidad Torcuato Di Tella, Almirante Juan Saenz Valiente 1010, C1428BIJ Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 24 March 2014

Keywords:
Perceptual decisions
Metacognition
Signal-detection theory
Sensory reliability
Introspection

ABSTRACT

Confidence in a perceptual decision is a judgment about the quality of the sensory evidence. The quality of the evidence depends not only on its strength ('signal') but critically on its reliability ('noise'), but the separate contribution of these quantities to the formation of confidence judgments has not been investigated before in the context of perceptual decisions. We studied subjective confidence reports in a multi-element perceptual task where evidence strength and reliability could be manipulated independently. Our results reveal a confidence paradox: confidence is higher for stimuli of lower reliability that are associated with a lower accuracy. We show that the subjects' overconfidence in trials with unreliable evidence is caused by a reduced sensitivity to stimulus variability. Our results bridge between the investigation of miss-attributions of confidence in behavioral economics and the domain of simple perceptual decisions amenable to neuroscience research.

1. Introduction

Subjective confidence is used ubiquitously in approximate everyday expressions such as – "I think…", "Maybe…", "I am sure that". In signal detection theory, confidence has a precise mathematical definition, indexing the probability that the decision is actually correct (Kepecs & Mainen, 2012). A large corpus of data suggests that the brain can be close to optimal when performing perceptual inferences under uncertainty, integrating multiple sources of evidence weighted by their reliability (Knill & Pouget, 2004; Körding & Wolpert, 2006). The view that confidence is an accurate index of decision uncertainty has been a conceptual anchor for neurophysiological and psychophysical studies of confidence (Bach & Dolan, 2012). Indeed, recent studies have demonstrated how cortical circuits can represent uncertainty in perceptual decisions and how these representations guide action (Kepecs, Uchida, Zariwala, & Mainen, 2008; Kiani & Shadlen, 2009).

The emerging picture of probabilistic inference in neuroscience, however, is in striking tension with the principles of behavioral economics. Decades of experimentation with "real-life" decision problems have shown that humans are

* Corresponding author. Address: Howard Hughes Medical Institute, Department of Neuroscience and Kavli Institute for Brain Science, Columbia University New York, 630 West 168th Street, P&S Building Room 16-409, New York, NY 10032, USA. Fax: +1 212 305 9608 4510.

E-mail address: ariel.zylberberg@gmail.com (A. Zylberberg).

http://dx.doi.org/10.1016/j.concog.2014.05.012
1053-8100/

sub-optimal decision makers and, specifically, that their estimation of confidence deviates from the predictions of probability theory (Griffin & Tversky, 1992; Lichtenstein, Fischhoff, & Phillips, 1977; Slovic, 1972; Slovic & Lichtenstein, 1968; Tversky & Kahneman, 1974, 1981). In economics, humans exhibit reliable inconsistencies: they rely on sub-samples of the data, focus on tokens (representative exemplars), ignore the variance (or reliability) of the distribution, and over-weight evidence confirming their previous commitments and choices (Griffin & Tversky, 1992; Tversky & Kahneman, 1973, 1974, 1981). Consider the following example of variance neglect from behavioral economics (Kahneman, 2011). When people are presented with the sentence "In a telephone poll of 30 seniors, 60% support the president", many conclude that "most seniors support the president". This leap of confidence ignores the fact that the reliability of a poll depends on the number of participants.

The present work is an effort to reconcile theories of probabilistic inference in neuroscience with behavioral economics (Griffin & Tversky, 1992; Jarvstad, Hahn, Rushton, & Warren, 2013; Wu, Delgado, & Maloney, 2009). We investigate if subjects are able to take the variance of visual stimuli into account and if conditions exist where the perceptual choices of subjects deviate systematically from their confidence reports.

2. Results

We studied sensory decision making with a field of 60 oriented line segments. Orientations were sampled from a uniform distribution of mean orientation μ and width φ. We controlled orientation jitter across the bars (exemplars) by varying φ, which was randomly selected from a set of three values (φ_i, $i = 1, 2, 3$) (Fig. 1a). Participants reported with a single (four-choice) response whether the average orientation was right (clockwise, CW) or left (counter-clockwise, CCW) relative to the vertical axis and the confidence in their choice ('high' or 'low').

Stimuli with more orientation jitter should generate more variance in the internal representation of orientation (de Gardelle & Summerfield, 2011). We first measured the form of this intuitive relationship, because it provides insight into the mechanism by which participants estimate the mean orientation. As expected, participants' accuracy increased with signal strength ($p < 10^{-8}$, likelihood-ratio test; see methods) (Fig. 1b). In addition, the subjects' orientation sensitivity decreased with jitter φ (the interaction between $|\mu|$ and φ was significant, $p < 10^{-8}$; likelihood ratio test). To formally relate accuracy to the value of μ, we relied on signal detection theory (SDT) (Green & Sweets, 1966). Specifically, we estimated accuracy as $P_{corr}(\mu) = \int_{-\infty}^{|\mu|} N(0, \sigma)$ and determined the best-fitting internal noise level σ for each stimulus class. We obtained values of $\sigma = [0.045, 0.072, 0.11]$ radians for stimuli with low, intermediate and high orientation jitter, respectively (Fig. 1c). As expected (de Gardelle & Summerfield, 2011), the stimuli with wider distributions of orientations were associated with more internal noise.

The signal-detection theory (Green & Sweets, 1966) also provides a theoretical framework to understand how signal and noise jointly determine confidence. Confidence should depend on the ratio between signal and noise so that the criterion that separates high from low confidence trials should vary with orientation jitter (Fig. 1d). However, if subjects ignore stimulus jitter (φ) when estimating the internal noise level (σ), the criterion that separates high from low confidence trials should be fixed (Fig. 1e). For simplicity, henceforth we refer to these two alternative models as *signal–noise* and *signal* models, respectively. When cast in the framework of the telephone poll, the *signal* model predicts equal confidence in polls with 30 and 300 seniors reporting that 60% support the president, whereas the *signal–noise* model predicts higher confidence in the larger pole. Note that the *signal* model mirrors the reliability misperception (Griffin & Tversky, 1992) from behavioral economics.

To distinguish between the *signal* and the *signal–noise* model, we estimated the patterns of confidence for different values of μ and the three empirically determined values of σ_i (Fig. 1b and c). For the *signal–noise* model, a trial was considered of high confidence whenever the random sample (x) exceeded a particular threshold that increased linearly with the orientation jitter (Fig. 1d), whereas the threshold did not depend on jitter in the *signal* model (Fig. 1e). These two models generate qualitatively different predictions. In the *signal–noise* model, confidence decreases if the jitter is higher (Fig. 1f). This pattern is intuitive: subjects are less confident for stimuli sampled from broad distributions. Conversely, in the *signal* model average confidence of the subjects is actually predicted to *increase* with jitter if the signal (μ) is small (Fig. 1g). This less intuitive dependence can be understood by observing that as σ increases, the fraction of the distribution crossing this fixed criterion increases and hence the probability that a random sample is assigned high confidence increases too.

We therefore analyzed how confidence depended on the signal and the noise (Fig. 1h). The confidence of the subjects tended to be higher for trials with a larger orientation jitter. To determine significance, we measured the influence of μ and φ on confidence with a logistic regression model. As predicted by *signal* and *signal–noise* models, the proportion of high-confidence responses increased with signal strength $|\mu|$ ($p < 10^{-8}$; likelihood ratio test). The influence of $|\mu|$ on confidence was weaker if orientation jitter was higher, as can be seen in Fig. 1h where the curves become flatter for larger orientation jitter (the combined effect of $|\mu|$ and φ on confidence was evaluated including a term $|\mu| \times \varphi$ in the regression model; this interaction had a significantly negative effect on the proportion of high-confidence responses ($p < 10^{-8}$; likelihood ratio test), which indicates that higher values of φ reduce the sensitivity of confidence to changes in $|\mu|$). Jitter φ had a main effect on confidence: larger jitter leads to more confident responses ($p < 10^{-8}$). At first sight, these results support the *signal* and are incompatible with the *signal–noise* model (Fig. 1h).

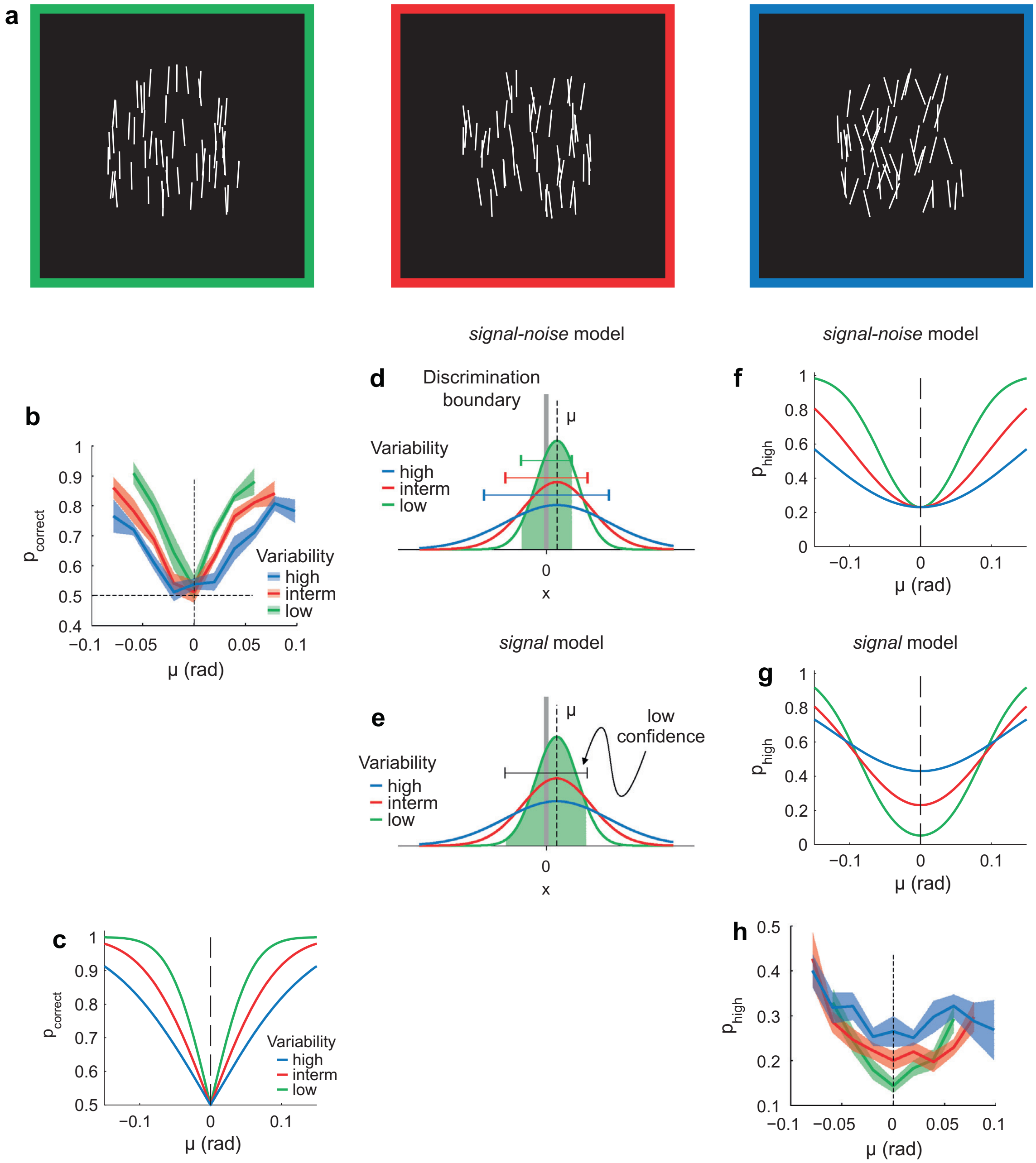

Fig. 1. Dissociable effects of signal strength and reliability on accuracy and confidence. (a) Subjects had to determine whether the average orientation of the lines was clockwise (CW) or counter-clockwise (CCW) relative to the vertical orientation. Line orientations were sampled from one of three uniform distributions of different widths. The panel shows example stimuli of low (green frame), intermediate (red) and high (blue) variability. Colored frames indicate jitter level but were not shown to the participants. (b) Proportion of correct choices as a function of average orientation (μ) averaged across participants. Shaded areas indicate s.e.m. (c) Proportion of correct choices as a function of μ for a signal-detection model where the internal noise (σ) of each distribution is obtained by a fit to the data in panel b. (d) A stimulus with orientation μ is represented as a sample (x) from a Gaussian distribution. Three distributions of different variances are shown. The colored horizontal lines indicate the range of x for which low-confidence responses should be given, if confidence depends on jitter. For illustration purposes, we set the criterion separating high- and low- confidence trials at 1.2 times the internal noise. Shaded green area highlights the low-confidence range for the low jitter condition. (e) As in panel d, but with a single confidence criterion for all jitter levels (Gorea & Sagi, 2000). Notice that if μ is held fixed, the proportion of low-confidence reports is larger for stimuli with lower jitter. (f) Proportion of high-confidence responses as a function of μ for the *signal–noise* model. Color codes are the same as in the previous panels. (g) As in panel f, for the *signal* model. (h) Proportion of high-confidence responses as a function of μ averaged across participants. Shaded areas indicate s.e.m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The *signal–noise* and *signal* models represent two extremes on a continuum of possible models for how strongly the confidence of subjects depends on stimulus variance. It seems unlikely that all our participants were completely blind to the orientation jitter. We therefore formulated a Bayesian model to quantify the dependence of each participant's confidence on stimulus variance (Fig. 2a).

The normative view (Gold & Shadlen, 2007; Wald & Wolfowitz, 1948) holds that subjects base their decisions on the relative probability of the two alternatives, known as the log posterior ratio *logPR*. According to Bayes' rule, this ratio depends on the log likelihood ratio and the priors. The contribution of the prior is zero if, as in our experiment, both orientations are equally probable and hence

$$logPR = \log\frac{p(\mathrm{CW}|x,\sigma)}{p(\mathrm{CCW}|x,\sigma)} = \log\frac{p(x|\mathrm{CW},\sigma)}{p(x|\mathrm{CCW},\sigma)} \tag{1}$$

where x and σ are the mean and variability of the internal tilt distribution, respectively. The choice of a rational agent depends on the sign of *logPR* (CW if $logPR > 0$) and its confidence should be an increasing function of the absolute value of the log posterior ratio $|logPR|$.

Eq. (1) shows that the likelihood depends c noise level σ, which, in turn, depends on the actual orientation jitter φ. Inaccurate estimates of σ would lead to inaccurate estimates of *logPR*, even when its calculation follows Eq. (1). For instance, if the same intermediate value of σ is used regardless of the actual jitter φ, large values of x would be taken as strong support for the CW hypothesis, ignoring that large jitter by itself increase the probability of observing large values of x. To allow for these inaccuracies in the estimation of *logPR* (and therefore confidence), we considered that the internal noise used to estimate confidence (σ_{Conf}) may differ from the internal noise directly derived from choice accuracy (σ_{Acc}). We estimated the participants' sensitivity to changes in jitter with a single parameter w:

$$\sigma^{\varphi}_{Conf} = w\sigma^{\varphi}_{Acc} + (1-w)\langle\sigma^{\varphi}_{Acc}\rangle_{\varphi} \tag{2}$$

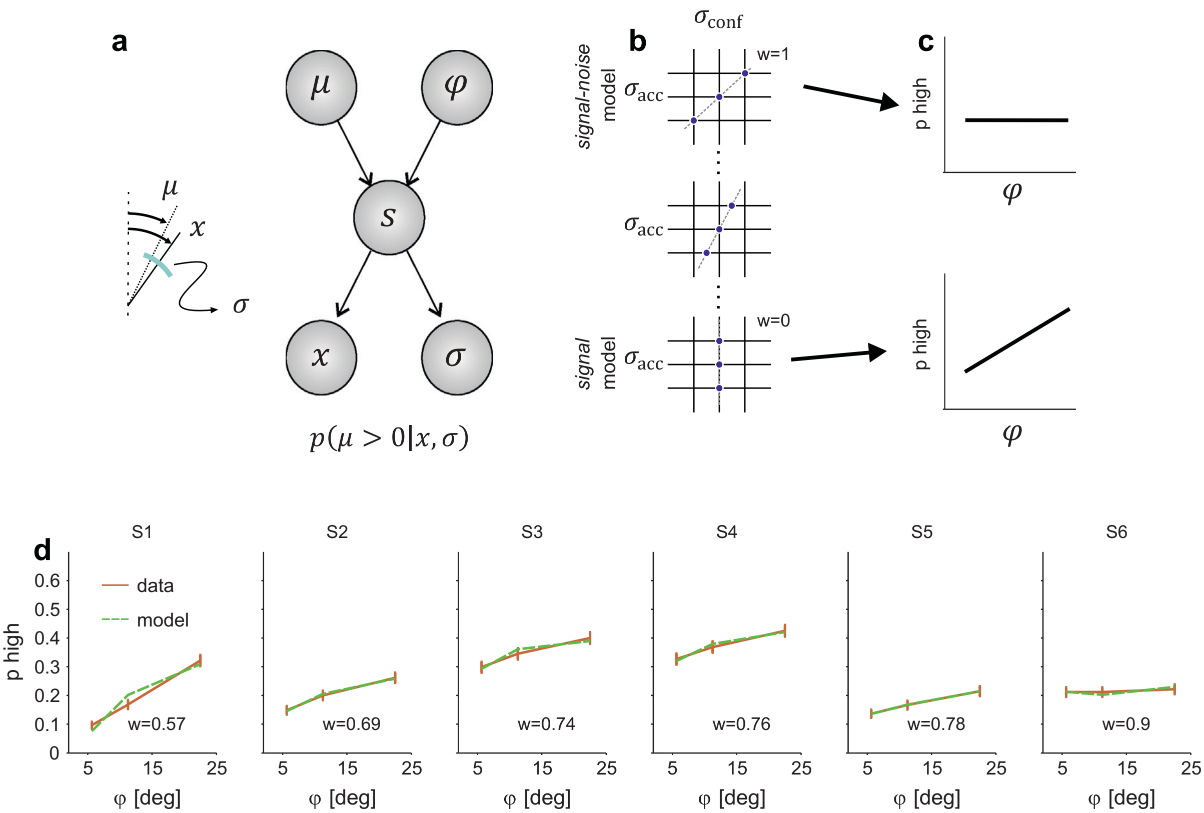

Fig. 2. Bayesian model of confidence judgments (a) A graphical model depiction of the stimulus generation and inference processes. The stimulus s is constructed sampling from a uniform distribution of parameters μ and φ. Subjects must infer the sign of μ given two noisy observations: a sample orientation x and an estimate of the variability of the internal representation of tilt σ. (b) A single parameter w is used to determine the influence of the internal noise level σ_{Acc} that determines accuracy as a function of μ, on the noise level σ_{Conf} that determines confidence. (c) The *signal* model of Fig. 1 corresponds to $w = 0$ and the *signal–noise* model to $w = 1$. When (as in our experiment) performance is held fixed for each jitter level (φ), jitter has no influence on the proportion of a high-confidence responses if $w = 1$. If $w = 0$ the proportion of high-confidence responses increases for stimuli with higher orientation jitter. (d) Proportion of high confidence responses as a function of φ (the variability of the distribution used to sample the orientation of each bar) in each of the participants. Data is shown in orange, and model fits in green. Error bars indicate s.e.m. The best-fitting w is indicated in each panel. For every subject, only a fraction of the variability which drives accuracy (σ_{Acc}) is available for the confidence report (σ_{Conf}).

Here σ_{Acc}^{φ} is the standard error of the internal distribution of tilt for jitter φ. It is obtained by fitting the accuracy data, and it is independent of confidence reports. Its average over φ is indicated as $\langle\cdot\rangle_{\varphi}$ The sensitivity of σ_{Conf} to changes in σ_{Acc} depends on a single parameter w, where $w = 1$ corresponds to the case where the internal noise estimate that determines confidence (σ_{Conf}) equals the internal noise level that determines the subjects' accuracy (σ_{Acc}) and $w = 0$ corresponds to a situation where σ_{Conf} is independent of orientation jitter (i.e. variance blindness; Fig. 2b and c). Fitting the model entails fitting the value of w to the proportion of high-confidence responses observed at each level of stimulus jitter (Fig. 2d).

The model revealed that subjects indeed underestimated changes in orientation jitter across trials, with values of w ranging between 0.57 and 0.9 (mean ± std: 0.74 ± 0.04) (Fig. 2d). At the same time, w was significantly larger than 0, which demonstrates that the subjects were not completely blind to the variation in jitter across trials either. It is of interest that this model explains the confidence paradox: a moderate underestimation of stimulus jitter can cause a profound overconfidence for stimuli with a high jitter (Fig. 1h).

In the first experiment the participants only received feedback about accuracy and not about the quality of their confidence report. Hence, a possible concern is that they may have been unmotivated to make accurate confidence reports and had no opportunity to calibrate their confidence judgments based on feedback. Furthermore, subjects only reported their confidence categorically as "high" or "low", which may have reduced the fidelity with which they convey their inner beliefs about certainty. To investigate whether confidence distortions persist when these concerns are alleviated, we conducted a second experiment. In our second experiment the subjects reported confidence on a continuous scale and received feedback that depended on their choice as well as on the reported confidence (see methods and Supplementary Fig. S1). Specifically, the subjects received positive confidence feedback if the reported confidence matched their accuracy.

Experiment two replicated our first experiment, because a larger jitter was associated with more confident responses, in particular if the signal strength μ was close to zero (Fig. 3). The influence of μ and φ on confidence was similar to that observed in the previous experiment, as revealed by a linear regression model. Confidence increased with signal strength $|\mu|$ ($p < 10^{-8}$; likelihood ratio test). The most important finding of the main experiment was also replicated: confidence increased with increasing stimulus jitter φ ($p < 10^{-8}$). Also in agreement with the main experiment, the influence of $|\mu|$ on confidence was weaker for higher orientation jitter (i.e., confidence curves become flatter for larger orientation jitter; Fig. 3). These results support the conclusions of our main experiment, showing that neither calibrating confidence through feedback nor making reports on a continuous scale cancel the positive influence of stimulus variability on confidence judgments.

3. Discussion

Confidence in a perceptual decision represents an estimate of the quality of an internal representation of sensory evidence, and should depend on the sensory evidence ('signal') and its reliability ('noise') (Yeung & Summerfield, 2012). Here we showed that the confidence of subjects in their perceptual decision deviates systematically from the veridical signal-to-noise ratio. Paradoxically, subjects assign higher confidence to stimuli with a larger variability that is associated with poorer objective performance. This is a reliable effect that we replicated in two experiments and that also occurred if the feedback explicitly encouraged accurate confidence reports. We showed how this result can be explained with a signal detection model where subjects do not optimally adjust their estimate of stimulus reliability. Our results demonstrate that subjects are not blind to variations in orientation jitter, but that they moderately underestimate the variance, which results in a relatively large distortion of their confidence.

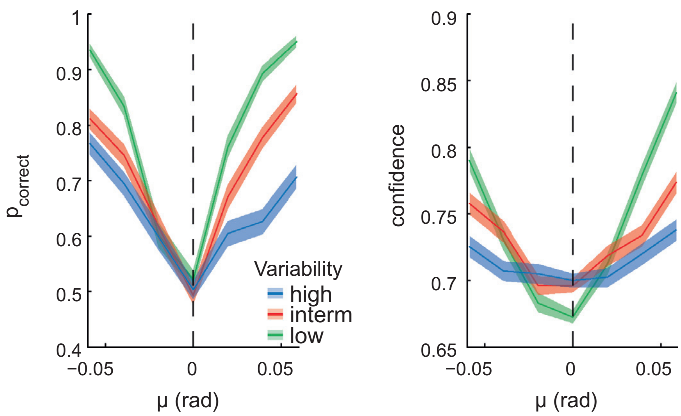

Fig. 3. Feedback does not abolish the illusion of confidence. Accuracy (left) and average confidence (right) for each level of signal (μ) and orientation jitter. Positive (negative) values of μ correspond to CW (CCW) tilt. Both accuracy and confidence increase with signal strength (tilt $|\mu|$) and are modulated by orientation jitter. Supplementary Fig. S1 illustrates the design of Experiment 2.

To assign an accurate degree of confidence to a decision, subjects must use a confidence criterion on the decision axis that depends on the reliability of the evidence. Instead, observers used a criterion that was biased toward the mean criterion across trials, which may reflect an inability to maintain multiple accurate decision criteria. Indeed, previous studies (Gorea & Sagi, 2000) demonstrated that observers tend to use a single decision criterion, even if they have to report about different stimuli. In addition, our findings are consistent with previous paradoxical results showing that directing attention away from a stimulus (Rahnev, Maniscalco, Luber, Lau, & Lisanby, 2011b; Rahnev et al., 2011a) or the application of TMS over the visual cortex may increase subjective visibility (Rahnev et al., 2011a,b). The absence of attention and the application of TMS pulses may cause a wider distribution of the internal decision variable and thereby increase confidence, similar to our *signal* model described above (Rahnev et al., 2011a,b).

Our results showing that people insufficiently adjust for variance in a simple perceptual decision parallel previous results from cognitive sciences and behavioral economics (Griffin & Tversky, 1992). In tasks probing people's ability to respond to general knowledge questions, subjective probability distributions tend to be narrower than warranted by their accuracy (see (Lichtenstein et al., 1977) for a review). For example, Pitz (1974) asked subjects to estimate the population of different countries, reporting the central tertile of the subjective distribution. Across subjects, only 16% (instead of one third) of the true values fell within this range. Fully neglecting the variance of the evidence, however, is not warranted by our results nor by previous studies showing that the more variable the evidence supporting an hypothesis, the more resistent it becomes to disconfirming evidence (Nassar, Wilson, Heasly, & Gold, 2010; Rehder & Hastie, 1996).

While our results are aligned with previous results from behavioral economics and cognitive sciences, they contrast with the emerging picture in neuroscience suggesting that the brain can be close to optimal in simple sensory decisions and in motor control (Knill & Pouget, 2004; Körding & Wolpert, 2006). It is generally considered that results and models from the cognitive domain do not extend to sensory decisions (Juslin & Olsson, 1997). However, recent studies comparing economic and perceptual decisions for carefully matched experimental designs suggest that the differences between economic and perceptual decisions might be more illusory that previously thought (Jarvstad et al., 2013; Wu et al., 2009). By showing that people partially neglect the reliability of the evidence when reporting confidence on simple perceptual decisions, the variance underweighting known from behavioral economics can now also be explored in the realm of sensory decision-making.

4. Experimental procedures

4.1. Experiments

Six subjects aged 18–26 participated in experiment one after giving informed consent. All had normal or corrected-to-normal vision and where naive to the purpose of the experiment. Each participant performed 6 blocks of 300 trials each.

Trials began with the presentation of a white fixation point (0.27°) on a black background. After 0.3 s, the oriented bars were presented at the center of the screen for 0.3 s. After stimulus offset, a blank screen was presented until the participant made a response. After the response, a colored dot (0.27°) presented for 0.1 s indicated whether the choice was correct (green) or incorrect (red). Trials were separated with a 0.4 s inter-trial interval during which the screen was blank. Responses were made with a standard keyboard. Choice and confidence were reported simultaneously. CCW (CW) decisions were communicated with the index and middle fingers of the left (right) hand, pressing keys *d* and *f* (*k* and *j*) for high and low confidence responses respectively.

The stimulus was composed of 60 bars uniformly arranged within a circular aperture of 6.65° All bars had a length of 1.34°. Orientations were sampled from a uniform distribution; the width φ of the distribution was pseudo-randomly selected on each trial from the set $\left[\frac{\pi}{16}, \frac{\pi}{8}, \frac{\pi}{4}\right]$. Trials from the three variance groups were randomly intermixed.

The mean orientation ($|\mu|$) was adjusted to maintain performance close to a value of 67%, with a 2-up 1-down staircase procedure. Independent staircases were conducted for each of the three variance groups. For the staircases, we adopted an initial value of $\frac{\pi}{20}$ and a step-size of $\frac{\pi}{160}$.

Twenty-five subjects performed the second experiment (aged 18–26), each completing 432 trials in blocks of 72 trials. In this experiment we used four fixed signal strengths which were the same for every participant: $|\mu| = \frac{\pi}{160} \times [0, 1, 2, 3]$ radians. We randomly selected 30 orientations and assigned each orientation to two bars, with opposite signs relative to the desired average tilt for the trial (μ). Thus, the sample mean was always equal to the true mean of the underlying distribution. Bar positions were sampled at random within the circular aperture. Participants had to maximize a numeric reward that was given after every trial, and which depended on both accuracy and confidence, as indicated next. While the main findings of the main experiment where replicated in the second experiment, confidence was less affected by sensory variability in the second experiment, indicating that the changes implemented were effective in increasing the calibration of confidence reports. Accordingly, more data was collected in the second experiment to assure the reliability of observed effects, which support the conclusions of our main experiment showing a positive influence of stimulus variability on confidence judgments.

4.2. Giving feedback on confidence reports

For the second experiment we adopted an operant definition of confidence, as is common in non-human investigations of confidence (Kepecs et al., 2008; Kiani & Shadlen, 2009). Subjects had to maximize a numeric reward, which depended on both accuracy and confidence. We designed a feedback function where the highest reward is gained if the reported confidence equals the objective performance. This function can be derived as follows. For a fixed level of performance, the average reward is the weighted average of the reward obtained in correct and error trials:

$$R = r_C \cdot p_{acc} + r_I \cdot (1 - p_{acc}), \tag{3}$$

where p_{acc} is the c, r_C is the reward obtained on correct trials, and r_I the reward obtained on incorrect trials.

If follows that r_C and r_I should depend on the subjective probability (p_{conf}) in a manner where reward is maximal if p_{conf} equals p_{acc}, for all values of p_{acc}. At this desired maximum, the derivative of p_{conf} to the average reward equals zero:

$$\frac{dR}{dp_{conf}} = \frac{r_C}{dp_{conf}} \cdot p_{acc} + \frac{r_I}{dp_{conf}} \cdot (1 - p_{acc}) = 0 \tag{4}$$

The solutions to this equation where $p_{acc} = p_{conf}$ have the following form for any integer n:

$$\begin{aligned} r_C &= \alpha \cdot \left(-p_{conf}^{n} + \tfrac{n}{n-1} \cdot p_{conf}^{n-1}\right) \\ r_I &= \alpha \cdot (-p_{conf}^{n}) \end{aligned} \tag{5}$$

We used Eq. (5) to determine the payoffs associated to correct and error trials (with $n = 3$ and $\alpha = 100$), and the shape of the resulting reward schedule has been illustrated in Supplementary Fig. S1 available online. It can be seen in this figure that the reward is indeed maximal if $p_{acc} = p_{conf}$, providing an incentive for accurate confident judgments.

4.3. Statistical analysis

The influence of μ and φ on accuracy and confidence was evaluated with logistic regression. For accuracy, the following model was fitted:

$$P_{correct} = [1 + exp(-\beta_0 - \beta_1|\mu| - \beta_2\varphi - \beta_3\varphi|\mu| - \beta_4 s)]^{-1} \tag{6}$$

where $|\mu|$ is the unsigned signal strength, φ is the width of the uniform distribution used to sample the oriented bars, s indexes the subject as a dummy variable, and the β's are the fitted coefficients. β_1 and β_2 determine the dependence of $P_{correct}$ on the signal and the noise, respectively, and β_3 indexes the interaction between $|\mu|$ and φ.

For confidence, the following model was fitted:

$$P_{high} = [1 + \exp(-\beta_0 - \beta_1|\mu| - \beta_2\varphi - \beta_3\varphi|\mu| - \beta_4 s - \beta_5 c]^{-1} \tag{7}$$

where the different variables are defined as above and c indicates whether the given response was correct or incorrect.

The statistical significance of the coefficients was evaluated with a likelihood ratio test, comparing nested regression models (with and without the regressors being evaluated) against a *chi-squared* distribution with a number of degrees of freedom equal to the difference in the number of independent variables between models (Hosmer & Lemeshow, 2000).

For the second experiment, we used a multiple linear regression model to accommodate the continuous scale of the confidence judgments, defined as follows:

$$Confidence = \beta_0 + \beta_1|\mu| + \beta_2\varphi + \beta_3\varphi|\mu| + \beta_4 s + \beta_5 c \tag{8}$$

where the parameters were defined as described above.

4.4. Model fitting

We implemented a Bayesian model to understand the impact of variance misperception in the construction of confidence. We used accuracy data to estimate the standard error of the internal distribution of evidence (σ_{Acc}), and then used the confidence reports to determine the degree to which confidence depended on this value of σ_{Acc}.

To derive the standard error of the internal evidence distribution (σ_{Acc}), we assumed that the likelihood function $p(x|\mu^{\varphi}, \sigma_{Acc}^{\varphi})$ was normally distributed. We estimated σ_{Acc}^{φ} for the different levels of jitter by setting the mean μ^{φ} to the orientation determined by the staircase procedure. We estimated σ_{Acc}^{φ} by matching the proportion of incorrect responses per participant and variability class (Green & Sweets, 1966), assuming that subjects selected option CW if $x > 0$ and option CCW otherwise (i.e., solved σ_{Acc}^{φ} such that $P_{error}^{\varphi} = \int_{-\infty}^{0} dx \cdot \Phi(x|\mu^{\varphi}, \sigma_{Acc}^{\varphi})$, where P_{error}^{φ} is the proportion of errors made at jitter level φ and Φ is the normal probability density function).

We fitted the proportion of high confidence responses as a function of the variability class (data shown in Fig. 2d), with w as a free parameter. The fits were based on 30,000 trials per participant. In each trial, we sampled x from a normal distribution, $x \sim \Phi(\mu^{\varphi}, \sigma_{Acc}^{\varphi})$, with φ selected randomly from the set of three possible values (the same values as for the experiment). For each trial, we computed the log of the posterior ratio as:

$$logPR = \log\frac{\Phi(x|\mu^{\varphi}, \sigma_{Conf}^{\varphi})}{\Phi(x| - \mu^{\varphi}, \sigma_{Conf}^{\varphi})} \tag{9}$$

with σ^{φ}_{Conf} computed as indicated in Eq. (2):

$$\sigma^{\varphi}_{Conf} = w\sigma^{\varphi}_{Acc} + (1 - w) \cdot S \quad (10)$$

where we set S to $\langle\sigma^{\varphi}_{Acc}\rangle_{\varphi}$. Different values of S are also capable of fitting the data of Fig. 2d as long as S is independent of stimulus jitter, i.e., as long as σ^{φ}_{Conf} becomes independent of φ when w equals 0.

We set a threshold over the absolute value of the log posterior odds ($|logPR|$) to classify trials as high- or low- confidence. This threshold was set by hand to match the total proportion of high confidence trials for model and data. We used matlab's *fminsearch* routine to search for the value of w that minimized the sum of squared differences between model and experiment for the data points shown in Fig. 2d. Results were identical if the threshold was also fit to the data.

Acknowledgments

This work was funded by CONICET, the Netherlands Organization for Scientific Research, the Human Frontier Science Program, and by the James McDonnell Foundation.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.concog.2014.05.012.

References

Bach, Dominik R., & Dolan, Raymond J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience, 13*(8), 572–586.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13341–13346.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 12380–12384.

Green, D., & Sweets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Krieger Publishing Company.

Griffin, Dale, & Tversky, Amos (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*(3), 411–435.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

Jarvstad, Andreas, Hahn, Ulrike, Rushton, Simon K., & Warren, Paul A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences, 110*(40), 16271–16276.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344.

Kahneman, Daniel (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1322–1337.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227–231.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759.

Knill, David C., & Pouget, Alexandre (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712–719.

Körding, Konrad P., & Wolpert, Daniel M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10*(7), 319–326.

Lichtenstein, Sarah, Fischhoff, Baruch, & Phillips, Lawrence D. (1977). *Calibration of probabilities: The state of the art*. Springer.

Nassar, Matthew R., Wilson, Robert C., Heasly, Benjamin., & Gold, Joshua I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience, 30*(37), 12366–12378.

Pitz, Gordon F. (1974). Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), *Knowledge and cognition*. New York: Wiley.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011a). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*(12), 1513–1515.

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2011b). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology, 107*(6), 1556–1563.

Rehder, Bob, & Hastie, Reid (1996). The moderating influence of variability on belief revision. *Psychonomic Bulletin and Review, 3*(4), 499–503.

Slovic, Paul (1972). From Shakespeare to Simon: Speculations - and some evidence - about man's ability to process information. *Oregon Research Institute Research Bulletin, 12*(2).

Slovic, Paul, & Lichtenstein, Sarah (1968). Importance of variance preferences in gambling decisions. *Journal of Experimental Psychology, 78*(4p1), 646.

Tversky, Amos, & Kahneman, Daniel (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

Tversky, Amos, & Kahneman, Daniel (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Tversky, Amos, & Kahneman, Daniel (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453–458.

Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics, 19*(3), 326–339.

Wu, Shih-Wei, Delgado, Mauricio R., & Maloney, Laurence T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences, 106*(15), 6088–6093.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1310–1321.