Cognition 234 (2023) 105377

Contents lists available at ScienceDirect

Cognition

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

The presence of irrelevant alternatives paradoxically increases confidence in perceptual decisions

Nicolás A. Comay [a,*], Gabriel Della Bella [a], Pedro Lamberti [b], Mariano Sigman [c,d], Guillermo Solovey [e,1], Pablo Barttfeld [a,1]

[a] Cognitive Science Group. Instituto de Investigaciones Psicológicas (IIPsi), CONICET-UNC, Facultad de Psicología, Universidad Nacional de Córdoba, Boulevard de la Reforma esquina Enfermera Gordillo, CP 5000 Córdoba, Argentina

[b] Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Córdoba, Argentina

[c] Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Buenos Aires, Argentina

[d] Facultad de Lenguas y Educación, Universidad Nebrija, Madrid, Spain

[e] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, UBA-CONICET, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Confidence
Perceptual decision making
Multiple alternatives
Bayesian confidence hypothesis
Computational modeling
Open data

ABSTRACT

Confidence in perceptual decisions is thought to reflect the probability of being correct. According to this view, confidence should be unaffected or minimally reduced by the presence of irrelevant alternatives. To test this prediction, we designed five experiments. In Experiment 1, participants had to identify the largest geometrical shape among two or three alternatives. In the three-alternative condition, one of the shapes was much smaller than the other two, being a clearly incorrect option. Counter-intuitively, confidence was higher when the irrelevant alternative was present, evidencing that confidence construction is more complex than previously thought. Four computational models were tested, only one of them accounting for the results. This model predicts that confidence increases monotonically with the number of irrelevant alternatives, a prediction we tested in Experiment 2. In Experiment 3, we evaluated whether this effect replicated in a categorical task, but we did not find supporting evidence. Experiments 4 and 5 allowed us to discard stimuli presentation time as a factor driving the effect. Our findings suggest that confidence models cannot ignore the effect of multiple, possibly irrelevant alternatives to build a thorough understanding of confidence.

1. Introduction

We constantly make decisions under uncertainty, accompanied by a feeling of confidence -a belief or a subjective feeling that our own thoughts, knowledge, performance, percepts or decisions are correct (Grimaldi, Lau, & Basso, 2015; Mamassian, 2016; Meyniel, Sigman, & Mainen, 2015). To be useful in our everyday activities, confidence has to reflect the true likelihood of being correct (Ais, Zylberberg, Barttfeld, & Sigman, 2016). However, this correspondence is not perfect, and several biases have been described in the literature (Rahnev & Denison, 2018), including overconfidence in difficult tasks and underconfidence in easy tasks (Baranski & Petrusic, 1994; Gigerenzer, Hoffrage, & Kleinbölting, 1991), a mismatch between confidence and accuracy on individual trials (Maniscalco & Lau, 2012), and overconfidence due to misperception of stimulus variability (Zylberberg, Roelfsema, & Sigman, 2014). However, despite these findings, the dominant view of the field is that confidence reflects the probability that the decision is correct. This definition is known as the Bayesian confidence hypothesis (BCH) (H.-H. Li & Ma, 2020; Meyniel et al., 2015; Sanders, Hangya, & Kepecs, 2016) and is inline with the classical assumption —known as independence of irrelevant alternatives (Luce, 1979)— that the subjective value of a stimulus is only dependent on the stimulus properties.

The Bayesian confidence hypothesis is based mostly on 2-alternative forced choice tasks (2AFC), because they allow for a relatively easy and straightforward computational modeling of confidence (Rahnev, 2020) using diverse approaches, such as signal detection theory (Fleming & Lau, 2014), accumulation of evidence (Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2013) and Bayesian modeling (Drugowitsch, Moreno-Bote, & Pouget, 2014; Galvin, Podd, Drga, & Whitmore, 2003; Kepecs & Mainen, 2012; Mamassian, 2016; Pouget, Drugowitsch, & Kepecs,

* Corresponding author.
E-mail address: nicocomay@gmail.com (N.A. Comay).

[1] Last authors.

https://doi.org/10.1016/j.cognition.2023.105377
Received 11 April 2022; Received in revised form 6 January 2023; Accepted 13 January 2023

2016). However, while methodologically useful, 2AFC tasks are a large oversimplification of the decision-making process, often implying more than two (or an undefined number of) options (H.-H. Li & Ma, 2020; Niwa & Ditterich, 2008; Rahnev, 2020). Including a larger number of alternatives may not only reveal that the methodological tools developed for 2AFC tasks are insufficient, but also that the proposed mechanisms are wrong (H.-H. Li & Ma, 2020). Specifically, the context in which we make decisions might alter our subjective interpretation of the facts (Ariely & Wallsten, 1995; Northoff & Mushiake, 2020). One example is the attraction effect (Huber, Payne, & Puto, 1982): when deciding between two alternatives (the focal options), the addition of a third one —similar to one of the focal alternatives regarding one attribute but inferior in another attribute— may increase the probability for choosing that focal option. Critically, this kind of context phenomena has been shown to play a role both in conceptual and perceptual processes (Trueblood, Brown, Heathcote, & Busemeyer, 2013), so they are thought to be a fundamental part in our decision processes. Moreover, it has recently been proposed that the allocation of attention —modulated by the stimuli location (Winter & Peters, 2021)— might also give rise to context influence on decision-making. Nevertheless, partly due to the prevalence of 2AFC tasks paradigms, the impact of these effects on confidence remains unknown.

Building on these ideas we explored whether a phenomenon already described in the literature for likelihood judgments (Windschitl & Chambers, 2004), line-up identifications (Charman, Wells, & Joy, 2011) and associative memory (Hanczakowski, Zawadzka, & Higham, 2014) takes place in multi-alternative perceptual decision-making. This effect is defined as "*the tendency to become more confident that a chosen response option is correct if it is surrounded by implausible response options*" (Charman et al., 2011, p. 479) and it is called the "dud-alternative effect". Bayesian models of confidence predict that confidence should remain virtually invariant when adding irrelevant alternatives, because these hardly affect the probability of being correct. Moreover, they practically do not affect the difference between the posterior probabilities that the two best options are correct, a recently proposed computation for confidence in multi-alternative tasks (H.-H. Li & Ma, 2020). On the other hand, the dud-alternative effect found in non-perceptual tasks predicts that confidence should paradoxically increase.

We investigated the influence of irrelevant alternatives on confidence in three perceptual decision making tasks. In contrast to other decision making processes, perceptual decision making tasks are mainly based on incoming sensory information. In Experiment 1 we studied the effect of a third, irrelevant stimulus on confidence and contrasted our data with predictions from four computational models. In Experiment 2 we replicated the results from Experiment 1 and, based on the models' simulations, extended them up to three irrelevant stimuli. In Experiment 3 we explored how stimuli-dependent our results are, using a task recently introduced to study confidence in a three alternative decision making context (H.-H. Li & Ma, 2020). In Experiments 4 and 5, we tested whether the differences found in the two previous paradigms are due to differences in the stimuli exposure duration.

2. Experiment 1

We designed a size discrimination task to evaluate the effect of the presence of a third, weak alternative (i.e. dud alternative) in confidence (Fig. 1). This task allowed us to manipulate the subjective value of the alternatives, making them more or less eligible only by varying their relative size.

2.1. Materials & methods

The task was programmed in JavaScript using the library jQuery. It ran on a JATOS (Lange, Kühn, & Filevich, 2015) server. The experiment could be run on mobile devices or desktop computers.

2.1.1. Participants

101 participants took part in Experiment 1 (63% females; mean age = 29.46, sd = 8.83). The study was approved by the ethical committee of the Institute of Psychological Research (IIPsi - CONICET - Córdoba, Argentina). Participants should read and accept an informed consent prior to the experiment. All participants reported no psychiatric, illegal drug consumption or neurological history. Instructions were written on screen prior to the experiment. Thirty four % of the sample performed the experiment on a mobile device, whereas the rest performed it on a computer.

2.1.2. Experimental design

The experiment consisted of a decision making task where participants had to decide which of two or three geometric shapes was the largest. It had 120 trials, 60 with two alternatives and 60 with three alternatives. The shapes were circles and squares, and the two largest shapes were always a circle and a square (to make the comparison harder). On half of the trials a circle was the correct option, and on the other half a square. The third alternative, when present, was randomly a square or a circle, and it was much smaller than the other 2 shapes in order to make it easily discarded (in fact, it was only chosen on 7 out of 11,751 trials). The size of the largest shape was sampled from a normal distribution with a mean value that depended on screen size (7500px^2 for a screen with a width size $\geq$ 600px; 6500px^2 for a screen with a width size <600px and $\geq$ 400; 5500px^2 for a screen with a width size <400px) and a standard deviation of one fourth of the mean value. The second

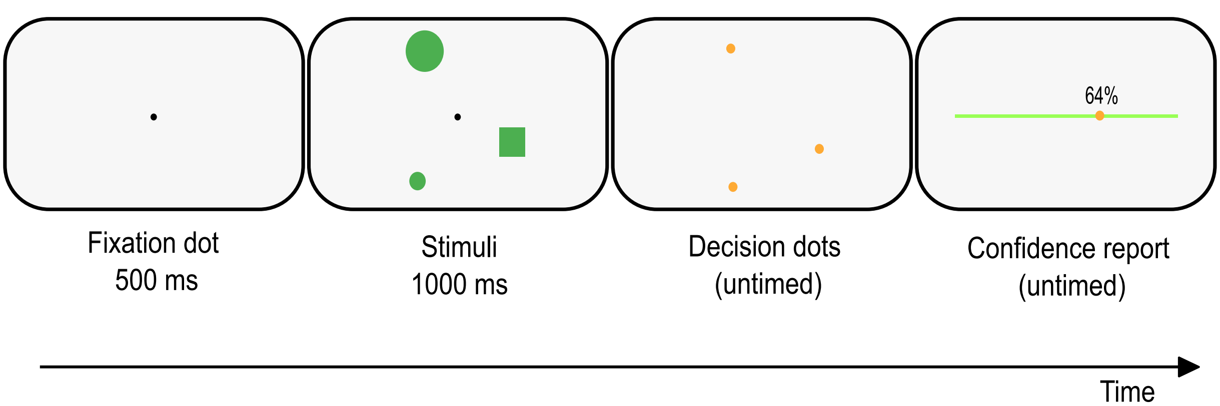

Fig. 1. Size discrimination task. Participants had to decide which was the largest shape. The size of the stimuli varied in every trial, and the relative size of the two largest shapes controlled the difficulty (established in 5 levels). Half of the trials had 2 alternatives and the other half had 3 alternatives, making up a total of 120 trials. The third alternative was very small when present, and varied in size (see Methods). After choosing the largest shape, participants had to report how confident they felt about their decision on a scale from 0% (not confident at all) to 100% (completely confident).

largest shape area was set to the 0.7, 0.8, 0.9, 0.93 or 0.95 of the area of the stimulus 1, and the third alternative' area (if present), was set to the 0.1, 0.2, 0.3, 0.4, 0.5 or 0.6 of the area of the stimulus 2. Consequently, there were 60 trials with 2 alternatives (of which there were 12 trials per each size of the second largest alternative) and 60 trials with 3 alternatives (of which there were 2 trials for each combination of size of the third and second largest alternative) (Table 1). Stimuli were randomly displayed in 3 equispaced positions (120 degrees) located in a circular array of a radius that varied with the screen size. The stimuli-array was randomly rotated so the positions changed across trials. Shapes were shown for 1 s, and the participants could respond by clicking the chosen shape or, once disappeared, by clicking a small circular dot marking the position. After that, participants had to report their confidence level on the decision on a scale from 0% (not confident at all) to 100% (completely confident). Participants had to move a dot that was initially hidden (until they moved the cursor) and appeared in the cursor's position. Specifically, in the confidence instructions we stated (in spanish in the actual experiment): "Once you have picked the shape that you think is the largest you will have to report your degree of confidence in that choice, that is, how sure you are that you picked the correct option" […] "the scale goes from 0% (COMPLETELY UNSURE, that implies that you don't have confidence at all in your decision of which was the largest shape) to 100% (COMPLETELY SURE, that implies that you are fully confident about your decision of which was the largest shape). YOU CAN USE INTERMEDIATE VALUES, the idea is that you report your confidence in your choice in the most precise way possible". On the first 3 trials we included, on the top of the screen, the question "How sure are you?"

2.1.3. Data analysis

We excluded 2 participants from subsequent analysis due to low overall performance (8 and 66% overall accuracy; the median performance for the sample was 85% with an interquartile range (Q3-Q1) of 6.7%). The definitive sample consisted of 99 participants. Our sample size was based on a pilot study with 20 participants in which we observed our main findings. We excluded trials with response times (RT) shorter than 200 ms and larger than 10 s. These exclusion criteria were not pre-registered, but including the discarded data does not modify the results. We conducted repeated measures ANOVAs and Tukey HSD post hoc tests (both with the software Statistica). Our predefined *p*-value for statistical significance was 0.05. We defined task difficulty as the ratio between the area of the two largest shapes. The ANOVA for incorrect trials was done with the highest levels of difficulty (0.9, 0.93, 0.95) because there were not enough incorrect trials in easier conditions (0.7 and 0.8).

2.2. Results

2.2.1. Response times

Task difficulty had a significant impact on the response times of both the decisions and the confidence reports. Higher difficulty was associated with larger RT ($F_{4,\ 392} = 130.95, p < 0.000001, \eta_p^2 = 0.57$ - decisions - Fig. 2a; $F_{4,\ 392} = 19.36, p < 0.000001, \eta_p^2 = 0.16$ - confidence reports - Fig. 2b). The presence of a dud alternative in half of the trials did not affect RT in none of the tasks ($F_{1,\ 98} = 0.79, p = 0.37, \eta_p^2 = 0.008$ - decisions - Fig. 2a; $F_{1,\ 98} = 0.002, p = 0.96, \eta_p^2 = 0.00002$ - confidence reports - Fig. 2b).

2.2.2. Performance

Participants' performance, as expected, decreased as task difficulty increased irrespective of the number of alternatives ($F_{4,\ 392} = 374.31, p < 0.000001, \eta_p^2 = 0.79$) (Fig. 2c). We found a marginally significant effect of the inclusion of the third alternative on performance, restricted to the middle difficulty ($F_{1,\ 98} = 4.04, p = 0.047, \eta_p^2 = 0.03$) (Fig. 2c). We didn't find a significant interaction between the number of alternatives and the task difficulty ($F_{4,\ 392} = 1.79, p = 0.12, \eta_p^2 = 0.017$) (Fig. 2c).

We also analyzed whether the shape of the third alternative had any effect on the decisions (i.e, a context effect regarding "square" and "circle" choices). In other words, we tested whether a squared (circled) third alternative made subjects more likely to choose a square (circle) as the largest figure. For trials where the chosen shape was a square, we found an interaction between the 2-alternative condition and the presence of the squared third alternative ($F_{4,\ 392} = 3.66, p = 0.006, \eta_p^2 = 0.04$). When comparing between 2 or 3-alternative conditions on the same level of difficulty, a Tukey HSD post-hoc test showed a significant difference only on the highest level of difficulty ($p = 0.02$). This means that participants were more likely to choose a square when the dud-alternative was also a square on that particular difficulty level. With respect to trials where the chosen shape was a circle, we again found an interaction between the 2-alternative condition and the presence of a circular third alternative ($F_{4,\ 392} = 2.72, p = 0.03, \eta_p^2 = 0.03$). However, a Tukey HSD post-hoc test showed no significant difference between the same difficulty conditions (i.e. participants were not more likely to choose a circle if the third alternative was a circle).

2.2.3. Confidence

Confidence, just like performance and as expected, decreased with increasing levels of task difficulty ($F_{4,\ 392} = 233.47, p < 0.000001, \eta_p^2 = 0.7$) (Fig. 3a). Importantly, we found a significant increment of confidence in the three-alternative condition ($F_{1,\ 98} = 33.25, p < 0.000001, \eta_p^2 = 0.25$) (Fig. 3a and b). Overall, 72 out of the 99 participants had an increased confidence in the three-alternative condition (Fig. 3c) as compared with the two-alternative condition. We also found a significant interaction between the difficulty of the task and the presence of a weak alternative ($F_{4,\ 392} = 6.85, p = 0.00002, \eta_p^2 = 0.06$) (Fig. 3a).

The effect of increased confidence is present both in correct and incorrect trials ($F_{1,\ 98} = 26.65, p = 0.000001, \eta_p^2 = 0.21$ - correct trials; $F_{1,\ 61} = 22.2, p = 0.00001, \eta_p^2 = 0.27$ - incorrect trials) (Fig. 4). Comparing the confidence increment between correct and incorrect trials (on the 3 highest levels of difficulty) no differences were found ($F_{1,\ 61} = 3.82, p < 0.055, \eta_p^2 = 0.06$). Regarding the size of the third alternative, we found a significant effect ($F_{6,\ 582} = 4.73, p = 0.0001, \eta_p^2 = 0.05$). A Tukey HSD post hoc test showed that the significant differences were between the 2 alternative condition and the conditions where the size of the dud alternative was 0.07 ($p = 0.048$), 0.21 ($p = 0.0007$), 0.28 ($p = 0.005$) and 0.35 ($p = 0.001$) of the area of the largest stimulus. This means that no differences were found across the third stimulus sizes (all differences are with respect to the 2 alternative condition). In other words, confidence increased irrespectively of the size of the dud-alternative (Fig. 3b).

We did not find an effect of the shape of the third alternative on confidence. That is, when considering if the dud alternative was a circle or a square only on the 3-alternative condition, no difference was found in confidence neither for trials where the chosen shape was a circle ($F_{1,\ 63} = 0.16, p = 0.69, \eta_p^2 = 0.003$) nor for trials where the chosen shape was a square ($F_{1,\ 83} = 1.69, p = 0.2, \eta_p^2 = 0.02$).

Table 1
Task structure.

Third alternative size (area3/area2)	Second alternative size (area2/area1)				
	0.7	0.8	0.9	0.93	0.95
Absent	12 trials	12 trials	12 trials	12 trials	12 trials
0.1	2 trials	2 trials	2 trials	2 trials	2 trials
0.2	2 trials	2 trials	2 trials	2 trials	2 trials
0.3	2 trials	2 trials	2 trials	2 trials	2 trials
0.4	2 trials	2 trials	2 trials	2 trials	2 trials
0.5	2 trials	2 trials	2 trials	2 trials	2 trials
0.6	2 trials	2 trials	2 trials	2 trials	2 trials

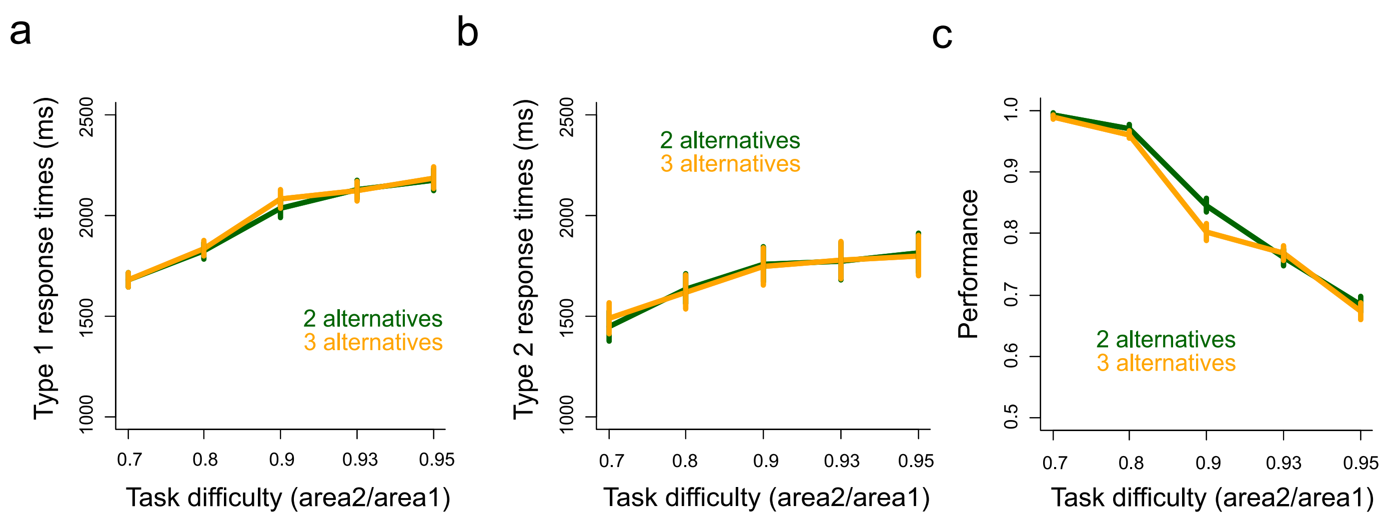

Fig. 2. The dud alternative did not have an effect on response times neither in (a) the decision task nor in (b) the confidence report. (c) The dud alternative only decreased task performance in difficulty 0.9. In all figures vertical bars refer to s.e.m. and “area2” and “area1” refer to the area of the second largest shape and the largest shape, respectively.

2.3. Discussion

The results of Experiment 1 extend the “dud-alternative” effect to a perceptual decision making task: confidence level increased with the addition of a small, clearly incorrect alternative. This effect is remarkably important because it implies that confidence does not directly reflect the probability of being correct, as stated by classical views (Meyniel et al., 2015; Sanders et al., 2016), and even deviates from recent models such as the difference of the probabilities of being correct of the two best options (H.-H. Li & Ma, 2020). According to these models, the addition of irrelevant alternatives should not affect confidence (because they hardly affect the probabilities of the rest of the alternatives), and, in any case, it should decrease it. However, the opposite pattern is observed in Experiment 1.

The dud-alternative had a small effect on performance. The effect is only restricted to a middle-level difficulty, where participants’ performance only decreased minimally. So it is likely that the task was not more difficult with the addition of the third alternative. The lack of effect in response times is consistent with this explanation, as more difficult tasks are expected to generate larger response times (Ratcliff, Smith, Brown, & McKoon, 2016). In consequence, the effect of the inclusion of the irrelevant alternative is mainly restricted to confidence.

The design of Experiment 1 allows for the possibility of testing a specific kind of “decoy effect”: the probability of choosing a geometrical shape as the largest one increases when the third alternative is of the same shape (Huber et al., 1982; Trueblood et al., 2013). However, we only found this effect for square choices, restricted to the highest level of difficulty. Moreover, this effect did not interact with the dud-alternative effect, suggesting that both are independent contextual effects.

Studies focused on the dud-alternative effect propose two main accounts for this phenomenon (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004).The first one is the “Average-residual” account: confidence reflects the difference between the evidence supporting the chosen option and the average of the evidence supporting the unchosen options (Eq. 3, see below). According to this idea, confidence increases with the addition of a dud alternative because its presence decreases the average of the evidence in favor of the unchosen options. The second proposed explanation for the dud effect is the “Contrast” account: confidence is obtained by a series of pairwise comparisons, where observers take the differences between the evidence supporting the chosen alternative and the evidence supporting each of the rest of the alternatives and sum them up (Eq. (4), see below). Confidence therefore increases with the presence of an irrelevant alternative, since there is another favorable comparison for the chosen option. We decided to formalize and test these verbal explanations in two computational models: the “Average-residual” model and the “Contrast” model. We compared these two models with the “Max” model –that states that confidence reflects the probability of being correct (Eq. (2), see below)– (H.-H. Li & Ma, 2020; Meyniel et al., 2015; Sanders et al., 2016) and the Difference model –that states that confidence reflects the difference between the probabilities of being correct of the two best alternatives (Eq. (1), see below)– (H.-H. Li & Ma, 2020).

3. Computational models

In order to evaluate whether the predictions of different models of confidence could fit our empirical results of Experiment 1 —and to propose some venues for future models— we carried out simulations for 4 models: the Difference model (H.-H. Li & Ma, 2020), the Max model (H.-H. Li & Ma, 2020; Meyniel et al., 2015; Sanders et al., 2016), the Contrast model (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004) and the Average-residual model (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004).

3.1. Materials & methods

3.1.1. Computational modeling

All models were developed under a Bayesian modeling approach (Ma, 2019). To set the generative model, we defined stimuli sizes as in Exp. 1. We simulated the internal responses on each trial. We assumed that observers make a noisy measurement x_i of each stimulus S_i. This measurement was modeled using a Gaussian distribution with mean on the actual value of the stimuli and standard deviation σ. Ideal observers compute the posterior probability that each stimuli is the largest, formally: $p_i = p(S_i = \max(S_1, S_2, S_3) \mid (x_1, x_2, x_3, s))$. As some variability in behavior is due to a “late noise” at the level of the decision variable (H.-H. Li & Ma, 2020), we assumed that observers do not maintain the exact posterior probability but a noisy version of it, q_i, modeled as a random sample of a Dirichlet distribution centered on the true posterior probabilities with a spread controlled by a parameter α (H.-H. Li & Ma, 2020). We assumed participants choose the stimulus with the highest $q(S_i \mid x_1, x_2, x_3, s)$. For readability, from now on we will refer to these noisy posterior probabilities as simply the posterior probabilities.

We simulated confidence levels according to four different models (Fig. 5a). For the Difference model (Diff), confidence level is the result of

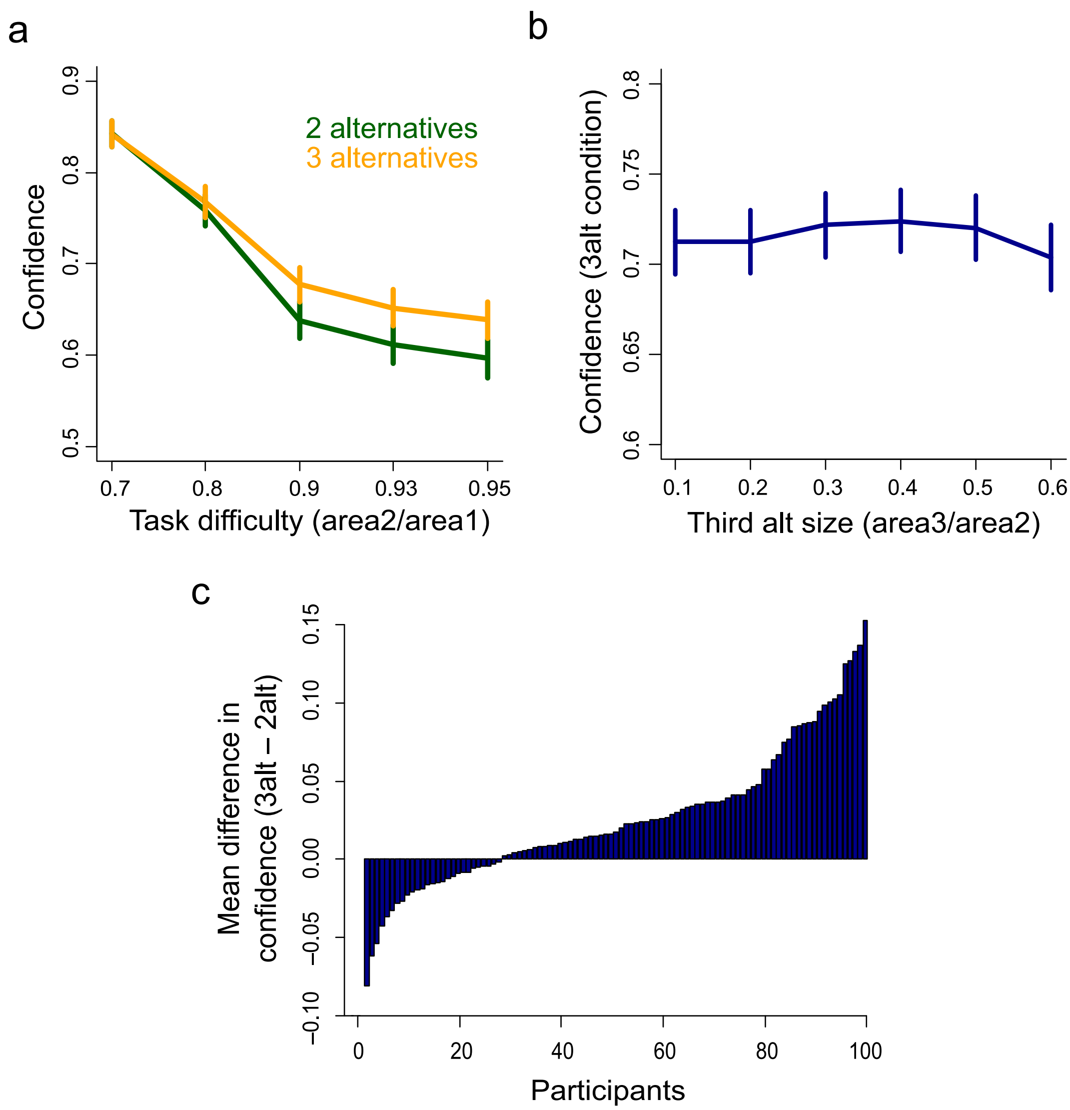

Fig. 3. (a) Confidence diminishes as a function of task difficulty but decreases less when a dud alternative is present (specifically at high levels of difficulty). (b) No effect was found regarding the size of the irrelevant alternative on confidence (i.e., confidence increased equally irrespectively of the third stimuli size). (c) Difference in confidence (2 vs 3 alternatives) per participant. Each line in (c) represents the change in confidence for each participant (considering all difficulty levels for each condition). A positive value means that confidence was higher when the dud alternative was present, whereas a negative value means that, conversely, confidence decreased in that condition.

the difference between the two highest posterior probabilities. In the Max model confidence level is the larger posterior probability. This model corresponds to the Bayesian Confidence hypothesis, stating that confidence reflects the probability that the decision is correct. The Contrast model and the Average-residual model are computational implementations of the two mechanisms proposed in the dud alternative literature to account for the effect (Charman et al., 2011; Windschitl & Chambers, 2004). The Contrast model states that confidence is obtained by a series of pairwise comparisons between the chosen option and the remaining. Following this, we modeled confidence by the sum of the differences between the largest posterior probability and all other posterior probabilities. The Average-residual model proposes that confidence reflects the evidence favoring the chosen option minus the average evidence of the remaining options. Therefore, we modeled confidence as the difference between the alternative with the highest posterior probability and the mean of the remaining posterior probabilities. Formally:

$$\text{Diff model}: \textit{confidence} = q_1 - q_2 \tag{1}$$

$$\text{Max model}: \textit{confidence} = q_1 \tag{2}$$

$$\text{Average} - \text{residual model}: \textit{confidence} = q_1 - \frac{(q_2 + q_3)}{2} \tag{3}$$

$$\text{Contrast Model}: \textit{confidence} = (q_1 - q_2) + (q_1 - q_3) \tag{4}$$

3.1.2. Model fitting

We fitted the models' free parameters separately for each participant's data (decisions and confidence) by maximizing the likelihood of the parameters. Formally, we maximized $L(\theta|\,data) = \ log\ p(data \mid \theta)$, where $L(\theta|\,data)$ is the log-likelihood of the parameters given the data, *data* is the decisions and the mean confidence level on each condition, and θ the parameters (σ, α, β_0 and β_1). The free parameters were sensory noise (σ), decision noise (α) and two parameters controlling the intercept (β_0) and slope (β_1) of the linear transformation that maps the confidence output of the model to the confidence data of the subject (using alternative mappings does not change the computational modeling results, see Supplementary material). We first fitted σ and α

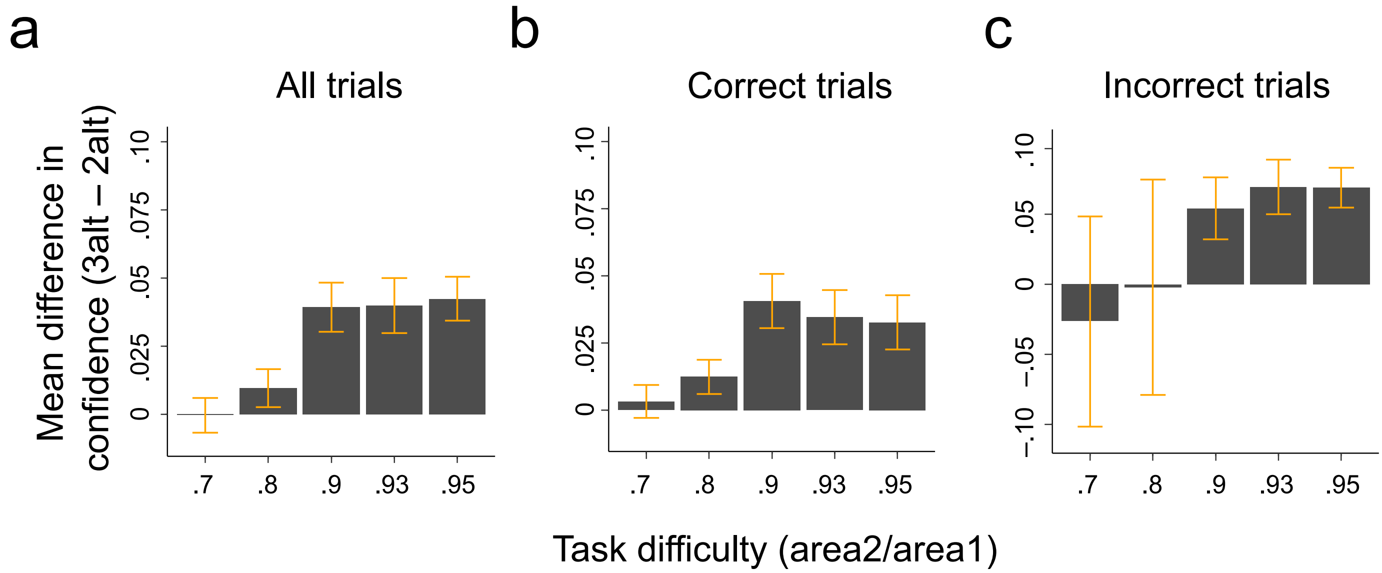

Fig. 4. Confidence increases when the third alternative is present. All panels depict the confidence difference between 3-alternative and 2-alternative conditions. The figure shows confidence difference for (a) all trials, (b) correct trials and (c) incorrect trials. Large s.e.m. in (c) at low difficulties (difficulties 0.7 and 0.8) are due to low numbers of incorrect responses.

using the participant's decisions and then used these two parameters as fixed to fit confidence levels. For that, we simulated 10,000 trials and calculated the proportion of trials where the model decision is equal to the decision of the participant. These numbers (one per trial) are an approximation of the data probability given the parameter values. We summed the log of these numbers, thus computing the log-likelihood of the parameters given the observed data. To find the best fitting parameter values we used the R function *optim* with the simulated annealing method to minimize the negative log-likelihood of the parameters' values given the observed data (which is equivalent to maximizing the likelihood of the parameters' values given the observed data). We then used the best σ and α values as fixed values for fitting β_0 and β_1. We simulated 100,000 trials per model taking the mean confidence by 1000 trials. Confidence data was divided by 100 to transform it to the scale of 0 to 1 and then rounded to 1 decimal. We fitted the model to the mean confidence level on each condition for each participant (fitting the whole trial distribution instead of the mean level did not qualitatively change our results, see Supplementary material). To fit confidence, we ran three optimization realizations with different starting points of the parameters, and kept the best estimation. Using the resulting parameters values we simulated 100,000 trials according to each model taking the mean confidence by 1000 trials and calculated their mean performance and confidence to compare these models' predictions to the participants' behavior. As all the models have the same number of free parameters we directly compared them using the summed log-likelihood across participants.

To evaluate the ability of our method to detect the true parameters for each subject, we performed parameter recovery by simulating, with the Average-residual model, the same amount of data that we had by participant (120 trials; 12 per condition) using values of σ, α, β_0 and β_1 sampled from our data and then fitting models to this simulated data with the mentioned procedure. The Pearson correlation between the true parameters and the estimated parameters was 0.83 for σ ($p < 0.0001$), 0.58 for α ($p < 0.0001$), 0.99 for β_0 ($p < 0.0001$) and 0.99 for β_1 ($p < 0.0001$). This suggests that our method successfully recovers the true parameters underlying each participant data.

3.2. Results

For both the Max model (summed log-likelihood = −7398.2) and the Diff model (summed log-likelihood = −7389) the presence of a third irrelevant option does not affect confidence (Fig. 5b). In contrast, for both the Contrast model (summed log-likelihood = −7686.03) and the Average-residual model (summed log-likelihood = −7365.97) confidence increases with the presence of an irrelevant option (Fig. 5b). However, the Contrast model shows a constant (difficulty-independent) confidence increase, whereas the Average-residual model accounts for the pattern shown in our data: the difference in confidence between three and two alternatives increases with the difficulty. Furthermore, if we extend simulations to a context with more dud alternatives, the Average-residual model predicts that confidence should increase monotonically with the number of dud alternatives (Fig. 5d).

3.3. Discussion

Among all tested models, both the Average-residual model and the Contrast model can recover the increment in confidence in the 3-alternative condition. However, only the Average-residual predicts a lower effect on easier trials. Contrary, both the Max model and the Diff model predict the opposite pattern: confidence remains the same when more alternatives are added into the decision context.

The Average-residual model predicts that confidence should be higher with more alternatives only when these alternatives are dud-alternatives, i.e. when the probability of these alternatives of being correct is close to zero (according to the model, if the extra alternatives are not dud-alternatives they will decrease the probability of being correct of the chosen option and then confidence will be lower than in 2-alternative case, see Supplementary Fig. 3a). This is because, with the addition of irrelevant alternatives, the probabilities of being correct of the competitive options remain the same while the average of the remaining options is lower. As a consequence, the model predicts that confidence increases monotonically with the number of dud-alternatives, since these alternatives progressively lower the average of the probabilities of being correct of the unchosen options (Supplementary Fig. 3b). We tested for this possibility in the next experiment.

4. Experiment 2

The Average-residual model predicts that confidence should increase monotonically with the number of dud-alternatives, because the

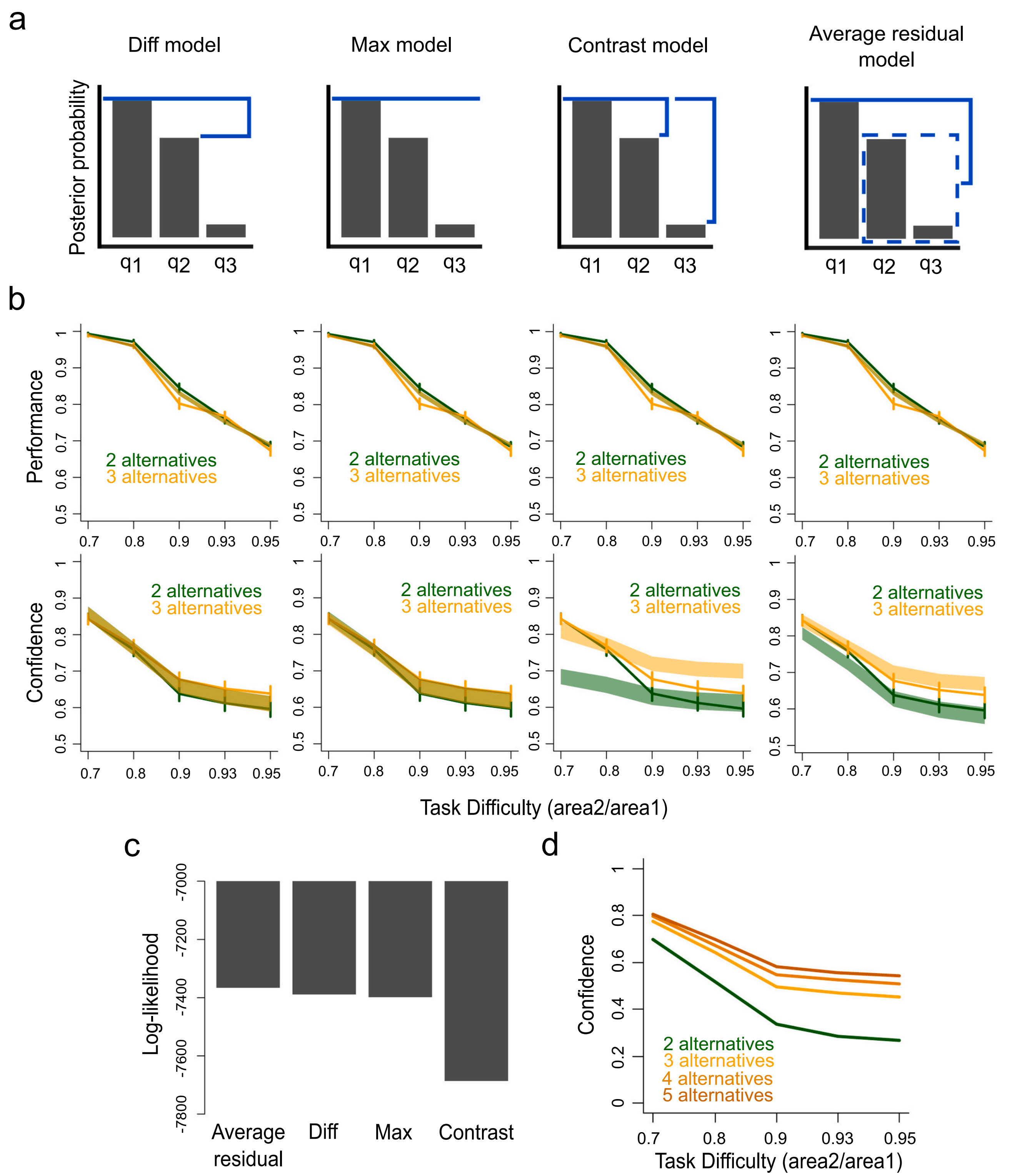

Fig. 5. (a) Sketches of confidence computation for each model. In the Average residual Model, confidence is obtained by the difference between the highest posterior probability and the mean of the remaining posterior probabilities. Confidence reflects the probability of being correct in the Max Model. The Difference Model states that confidence reflects the difference between the two best posterior probabilities. Finally, in the Contrast Model confidence is the result of the sum of the differences between the highest posterior probability and each of the remaining posterior probabilities. (b) Model fits for performance (first row) and confidence (second row). Both the max model and the Diff model show that confidence should remain invariant when a dud-alternative is added to the decision context. Conversely, the Average-residual model and the Contrast model show that confidence should be higher. However, only the Average-residual model replicates the pattern found in our experiment. Shaded regions are the standard error of the mean of model fits. (c) The log-likelihood of each model. The Average-residual model is the best fitting model regarding our data. (d) The Average-residual model predicts that confidence should increase with the number of dud alternatives, as these alternatives reduce even more the average of the remaining options.

presence of more irrelevant alternatives decreases the average of posterior probabilities of non-chosen alternatives with a minimal effect on the posterior probabilities of the competing alternatives (Fig. 5d). In this second experiment we not only aimed to replicate the results of Experiment 1 but to see whether confidence increases with the number of dud alternatives (Fig. 6).

4.1. Materials & methods

As in Experiment 1, the task was programmed in JavaScript using the library jQuery and conducted online on a JATOS (Lange et al., 2015) server. Experiment 2 involved more experimental conditions and, thus, more trials. To guarantee participants' attention during the task, the experimenter made a video call and remained available for the duration of the experiment.

4.1.1. Participants

18 participants took part in this experiment (62% females; mean age = 27.17, sd = 4.77). The study was approved by the ethical committee of the Institute of Psychological Research (IIPsi - CONICET - Córdoba, Argentina). Participants should read and accept an informed consent prior to the realization of the experiment. All participants reported no psychiatric, illegal drug consumption or neurological history. The tasks' instructions were written on screen prior to the experiments. All participants performed the task on a computer. Each participant received a payment of roughly 4 US dollars.

4.1.2. Experimental design

The design of Experiment 2 was similar to that of Experiment 1 but involving a larger number of conditions: 480 trials including 120 trials of 2, 3, 4 or 5 alternatives. As in Experiment 1, only two alternatives were large enough to compete for the correct answer; the others were much smaller in order to make them easily discarded (irrelevant alternatives). In trials where the irrelevant alternatives were present, they all had the same size, and each one was randomly a square or a circle. These dud alternatives also varied in size (as in Experiment 1). Participants had to report their confidence level in the decision in the same way as in Experiment 1.

4.1.3. Data analysis

No subject was excluded from the analysis. We followed a similar analysis strategy than in Exp 1 (including the excluded trials does not modify the results). We excluded the easiest level of the task from the performance analysis because one category (the one with 5 alternatives) had perfect performance. The dissociation between correct and incorrect trials carried out in Experiment 1 was not possible because there were not enough incorrect trials.

4.2. Results

4.2.1. Response times

Decisions RT increased with difficulty ($F_{4,\,68} = 23.47$, $p < 0.000001$, $\eta_p^2 = 0.58$) (Fig. 7a). This effect was not found in confidence reports RT ($F_{4,\,68} = 1.59$, $p = 0.18$, $\eta_p^2 = 0.08$) (Fig. 7b). The weak alternatives did not impact the RT on either of the tasks ($F_{3,\,51} = 0.26$, $p = 0.84$, $\eta_p^2 = 0.01$ - decisions - Fig. 7a; $F_{3,\,51} = 0.57$, $p = 0.63$, $\eta_p^2 = 0.03$ - confidence report - Fig. 7b).

4.2.2. Performance

Regarding task performance, participants' accuracy decreased as the difficulty increased ($F_{3,\,51} = 90.45$, $p < 0.000001$, $\eta_p^2 = 0.84$) (Fig. 7c). Contrary, the dud alternatives did not affect participants' performance ($F_{3,\,51} = 2.62$, $p = 0.06$, $\eta_p^2 = 0.13$) (Fig. 7c).

4.2.3. Confidence

Confidence level decreased with task difficulty ($F_{4,\,68} = 54.63$, $p < 0.000001$, $\eta_p^2 = 0.76$) (Fig. 8a). Replicating the main finding in Experiment 1 confidence increased with the presence of irrelevant alternatives ($F_{3,\,51} = 3.81$, $p = 0.01$, $\eta_p^2 = 0.18$) (Fig. 8a). We conducted a Tukey HSD post hoc analysis, showing that there was a significant difference between confidence with 2 and 5 alternatives ($p = 0.008$). To further explore this result, we conducted ANOVAs separately for each remaining conditions (3 and 4 alternatives vs 2 alternatives). We found a significant effect on confidence for the 3-alternative condition ($F_{1,\,17} = 9.06$, $p = 0.007$, $\eta_p^2 = 0.34$) not detected by the post hoc test. No interaction was found on this condition ($F_{4,\,68} = 0.5$, $p = 0.72$, $\eta_p^2 = 0.03$). For the condition with 4 alternatives we found a marginally significant effect ($F_{1,\,17} = 3.58$, $p = 0.075$, $\eta_p^2 = 0.17$) and a significant interaction between the amount of alternatives and the task difficulty ($F_{4,\,68} = 2.61$, $p = 0.04$, $\eta_p^2 = 0.13$). We did not replicate the interaction between difficulty and the irrelevant alternatives found in Experiment 1 ($F_{12,\,204} = 1.13$, $p = 0.33$, $\eta_p^2 = 0.06$) (Fig. 8a). Confidence was not affected by the number of dud-alternatives ($F_{2,\,34} = 0.95$, $p = 0.39$, $\eta_p^2 = 0.05$) (Fig. 8a and Fig. 8b). Thirteen out of eighteen participants showed the effect of increased confidence in the dud alternative conditions (Fig. 8c). There was a significant effect of the dud alternative size on confidence ($F_{6,\,102} = 2.49$, $p = 0.03$, $\eta_p^2 = 0.13$).

4.3. Discussion

In Experiment 2 we replicated the dud-alternative effect found in Experiment 1. However, we did not find evidence that adding more dud-

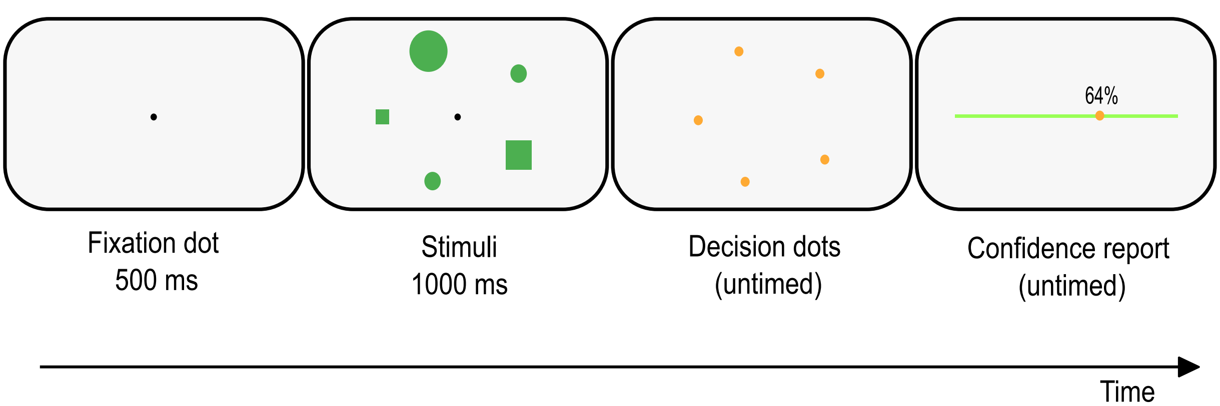

Fig. 6. Size discrimination task. Participants had to decide which one of the shapes was the largest. Stimuli size varied in every trial, and the relative sizes of the two largest shapes controlled the difficulty (established in 5 levels). There were 4 conditions: 2, 3, 4 and 5 alternatives, each one with 120 trials. After that decision, participants had to report their confidence level on a scale from 0% (not confident at all) to 100% (completely confident). Only two alternatives competed for the correct answer, the others (when present) were relatively small and easily discarded.

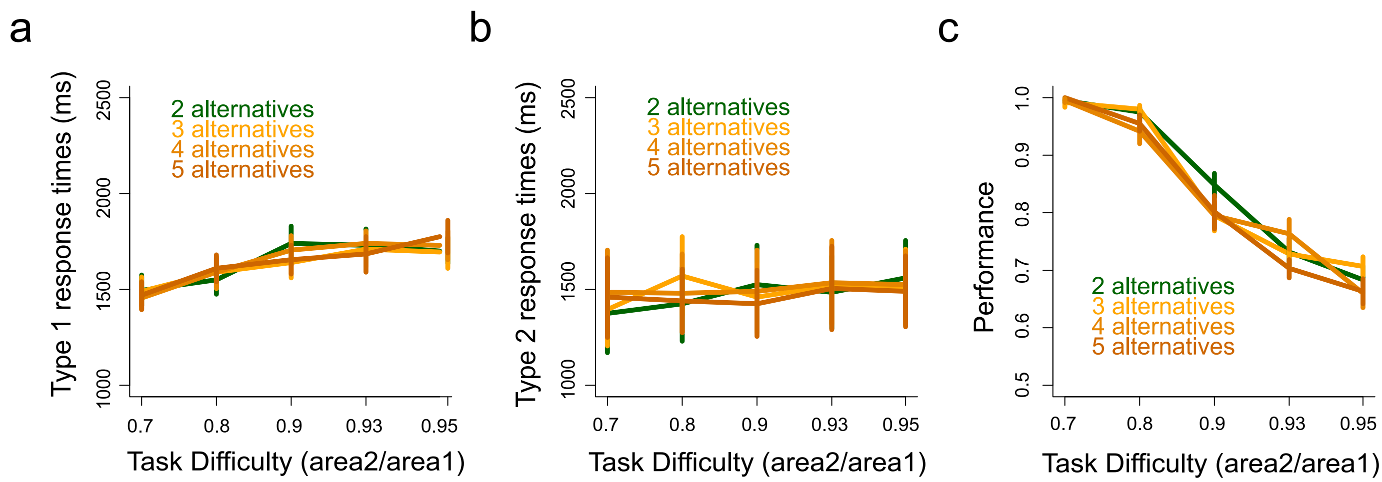

Fig. 7. As in Experiment 1, dud alternatives did not have an effect in RT, neither in (a) decisions or in (b) confidence reports. (c) Performance did not vary with the addition of dud alternatives.

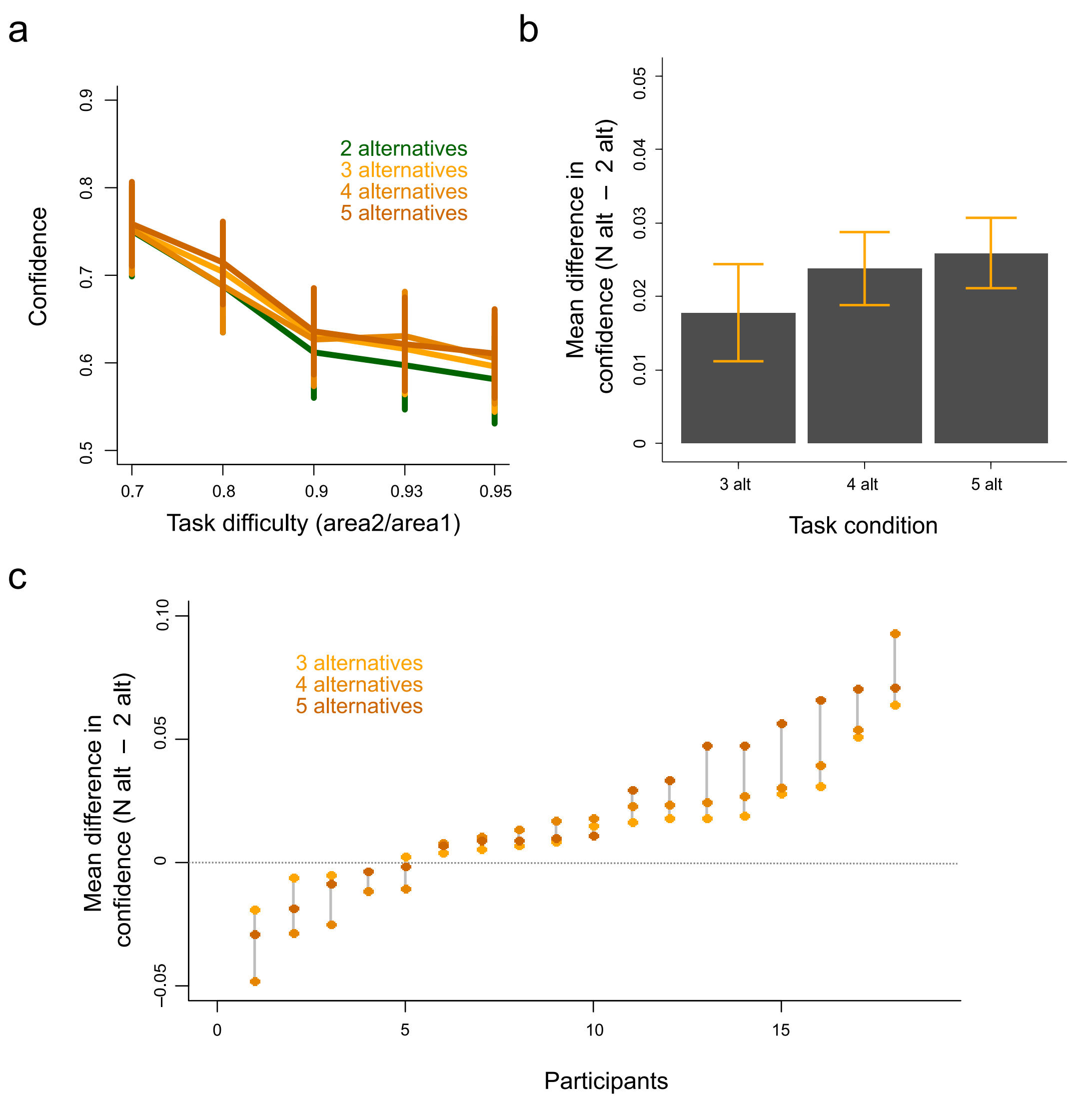

Fig. 8. (a) As in Experiment 1, confidence decreased with task difficulty but increased when weak alternatives were present. (b) Confidence difference between the 3, 4 and 5 alternative conditions and 2-alternative condition considering the highest levels of difficulty (0.9, 0.93 and 0.95). (c) Average difference in confidence by condition and participant.

alternatives influences confidence. Although this constitutes a challenge for the Average-residual model, one should keep in mind that the prediction of the model is that the increase in confidence becomes smaller with the addition of each dud alternative (because the average of the probabilities of the non-chosen options start to decrease less with each added alternative), making this effect, if it exists, much difficult to detect. Moreover, with higher levels of sensory or decision noise the model predicts no difference between conditions. Further research is needed to evaluate if the monotonic increase in confidence actually exists and if the Average-residual model can account for its presence or absence by fitting it to participants' behavior.

Interestingly, performance was not affected by the inclusion of the dud-alternatives. This confirms the notion that these alternatives are not making the task harder (despite the small but significant effect on performance found in Experiment 1).

We extended the dud-alternative effect to a perceptual decision making task in Experiments 1 and 2. However, we wanted to test if this effect was present in a different task, the one where the Diff model was developed (H.-H. Li & Ma, 2020), to evaluate whether the predictions of the Average-residual model hold even in the context where the Diff model was the best model. Therefore, we run Experiment 3.

5. Experiment 3

To evaluate whether the dud-alternative effect replicates in a different multi-alternative task, we adapted the target categorization task from Li and Ma (2020) (Fig. 9).

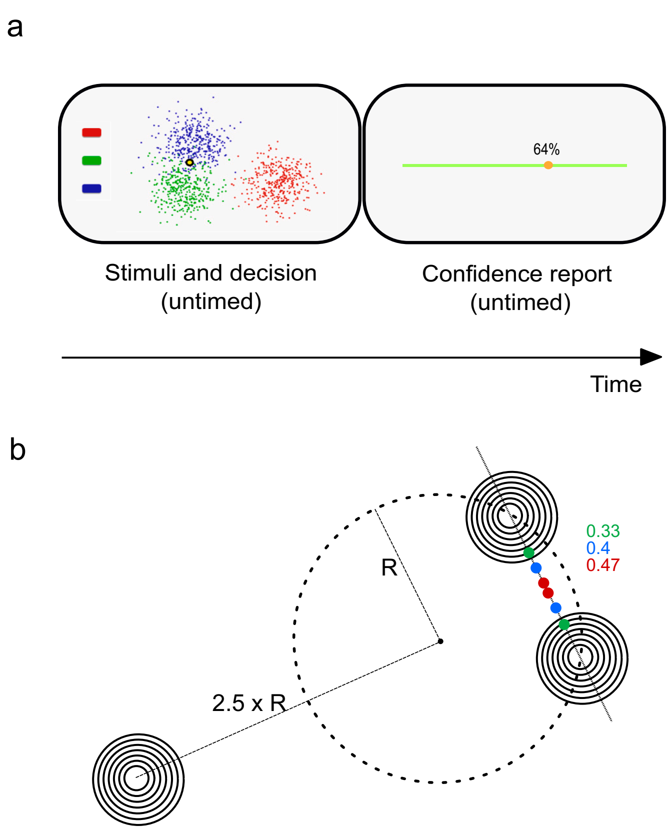

Fig. 9. Target categorization task. (a) Participants decided which cloud of dots the yellow target dot belonged to. Half of the trials had three alternatives and half two. When present, the third cloud was far from the target in order to make it ineligible, following the logic of the previous two studies (red cloud in the figure). After the decision, participants reported their confidence level on a continuous bar. (b) The underlying structure of the task (see Materials & methods). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.1. Materials & methods

As the two previous experiments, the third experiment was conducted online on JATOS (Lange et al., 2015) and programmed in JavaScript using the library jQuery.

5.1.1. Participants

63 participants took part in this experiment (62.5% females; mean age = 30.16, sd = 12.32). The study was approved by the ethical committee of the Institute of Psychological Research (IIPsi - CONICET - Córdoba, Argentina). Participants should read and accept an informed consent prior to the realization of the experiment. All participants reported no psychiatric, illegal drug consumption or neurological history. The tasks' instructions were written on screen prior to the experiment. All participants completed the experiment on a computer.

5.1.2. Experimental design

The experiment consisted of a decision making task where participants had to decide which cloud of dots a target dot belonged to. It had 120 trials, 60 with two alternatives and 60 with three alternatives (i.e., two or three clouds). Dots were normally distributed in space. The mean of the distributions were set in space such that the two main clouds were closer together. More specifically, the mean of these distributions were located in a circle of radius R centered at a fixation point and separated by an angle of 120 degrees. In trials where the third cloud was present, its mean was located at a distance 2.5xR from the fixation point and at an angle of 120 degrees from each of the mean locations of the other two clouds. Each dot cloud had 375 dots. Dot size and standard deviation of the distribution was responsively set, according to the presentation screen. The target dot was yellow with a black border. The target was located in a segment that begins and ends at the mean values of the 2 competing clouds distributions. The position of the target was parameterized by a "task difficulty" variable d that takes values from 0 to 0.5, where 0 represents one of the extremes and 0.5 represents the middle, equidistant from the mean values of the distributions. We used $d = \{0.33, 0.4, 0.47\}$ where 0.33 is the easiest and 0.47 the most difficult condition. See Fig. 9b for a graphical reference of the task. Stimuli remained on the screen until the participant made a decision, clicking on color coded screen buttons. Participants reported their confidence level on the decision on a continuous scale, as in Experiments 1 and 2. Similar to Experiments 1 and 2, the confidence instructions stated (spanish in the actual experiment): "your second task consists in reporting your degree of confidence in your choice, that is, how sure you are to have picked the correct option" […] "the scale ranges from 0% (COMPLETELY UNSURE, that implies that you don't have confidence at all in your decision to which cloud of dots did the yellow dot belonged to) to 100% (COMPLETELY SURE, that implies that you are fully confident about your decision to which cloud of dots did the yellow dot belonged to). YOU CAN USE INTERMEDIATE VALUES, the idea is that you report your confidence in your choice in the most precise way possible". On the first 3 trials we included, on the top of the screen, the question "How sure are you?"

5.1.3. Data analysis

We excluded 7 participants from the analysis (3 due to low performance -our accuracy cutoff point was an accuracy of 0.6-, and 4 due to lack of variability in the confidence report (confidence level at 100% in >90% of the trials). The final sample consisted of 52 participants (including all participants does not modify the results). We excluded trials with RT lower than 200 ms both for decisions and confidence reports. We excluded trials with a response time >20 s instead of 10 s as in our previous tasks since RTs are larger in this task. These exclusion criteria were not pre-registered, but including the discarded data does not modify the results. We conducted repeated measures ANOVAs and Tukey HSD post hoc tests (both with the Statistica software) to estimate the impact of the difficulty of the task and the dud alternatives in RT, performance and confidence. Our predefined *p*-value for statistical significance was 0.05.

5.2. Results

5.2.1. Response times

Task difficulty had a significant effect on decisions RT ($F_{2,\ 102} = 34.40$, $p < 0.000001$, $\eta_p^2 = 0.40$) (Fig. 11a). RT were also larger for the 3-alternative condition ($F_{1,\ 51} = 14.38$, $p = 0.0004$, $\eta_p^2 = 0.22$) (Fig. 10a). No interaction was found between the number of alternatives and task difficulty ($F_{2,\ 102} = 0.26$, $p = 0.77$, $\eta_p^2 = 0.005$) (Fig. 10a). Task difficulty affected confidence reports RT ($F_{2,\ 102} = 3.63$, $p = 0.03$, $\eta_p^2 = 0.07$). The number of alternatives did not impact confidence reports RT ($F_{1,\ 51} = 3.20$, $p = 0.08$, $\eta_p^2 = 0.059$). No interaction was found between task difficulty and the number of alternatives ($F_{2,\ 102} = 1.01$, $p = 0.37$, $\eta_p^2 = 0.02$).

5.2.2. Performance

Performance, as expected, decreased as the task became more difficult ($F_{2,\ 102} = 418.85$, $p < 0.000001$, $\eta_p^2 = 0.89$) (Fig. 10c). The presence of a third alternative did not have an effect on performance ($F_{1,\ 51} = 0.78$, $p = 0.38$, $\eta_p^2 = 0.015$) (Fig. 10c).

5.2.3. Confidence

Confidence, as expected and in line with performance, decreased when the task got more difficult ($F_{2,\ 102} = 97.59$, $p < 0.000001$, $\eta_p^2 = 0.66$) (Fig. 11a). In contrast with the results from Experiments 1 and 2, we did not find evidence for higher confidence in the 3-alternative condition ($F_{1,\ 51} = 0.485$, $p = 0.49$, $\eta_p^2 = 0.009$) (Fig. 11b) and there was not an interaction between the number of alternatives and task difficulty ($F_{2,\ 102} = 1.84$, $p < 0.16$, $\eta_p^2 = 0.035$).

We explored the impact of the presence of the third alternative on confidence level separately for correct and incorrect responses, as in Experiment 1 (Fig. 12). For correct trials only, confidence was not affected by the number of alternatives ($F_{1,\ 51} = 0.03$, $p = 0.86$, $\eta_p^2 = 0.0006$) and we did not find an interaction between task difficulty and the number of alternatives ($F_{2,\ 102} = 2.61$, $p = 0.078$, $\eta_p^2 = 0.05$). For incorrect trials, we did not find an effect of the amount of alternatives ($F_{1,\ 25} = 0.045$, $p = 0.83$, $\eta_p^2 = 0.002$) or an interaction between task difficulty and the number of alternatives ($F_{2,\ 50} = 1.40$, $p = 0.26$, $\eta_p^2 = 0.05$) (Fig. 12c).

5.3. Discussion

The target categorization task of Experiment 3 does not show a dud-alternative effect: confidence remained the same despite the addition of a third, clearly incorrect alternative. The inclusion of the dud-alternative did not affect performance or confidence, but made response times longer. This could be one of the reasons that this experiment bears no dud-alternative effect, since longer response times are usually associated with more difficult tasks and with lower confidence levels (Kiani, Corthell, & Shadlen, 2014). Another possible reason underlying the lack of the dud-alternative effect is the different presentation times of the stimuli: up to 1 s in experiments 1 and 2, and untimed in Experiment 3. In fact, the computations underlying confidence in purely perceptual tasks with brief stimuli presentations could be different because confidence has to rely more on the internal representation of the stimuli (Rahnev, 2020; Yeon & Rahnev, 2020). Consequently, we decided to test in Experiment 4 and 5 whether the dud-alternative effect is present both in the size discrimination task with unlimited presentation time of the stimuli (as in Experiment 3) and in the target categorization task with the stimuli presentation time restricted to one second (as in Experiments 1 and 2).

6. Experiment 4

To investigate whether stimuli presentation time is driving the dud alternative effect, we conducted two more experiments: Experiment 4 consisted of the same size discrimination task as in Experiment 1 but stimuli were present up to the time of the decision as in Experiment 3; Experiment 5 consisted of the same target categorization task as in Experiment 3 with stimuli being on screen for 1 s as in Experiments 1 and 2.

6.1. Materials & methods

As in the previous experiments, Experiment 4 was conducted online on JATOS (Lange et al., 2015) and programmed in JavaScript using the library jQuery.

6.1.1. Participants & experimental design

34 participants took part in this experiment (73.5% females; mean age = 25.5, sd = 6.38). One subject was excluded due to slow response times (see 6.1.2 Data analysis), so the final sample consisted of 33 participants.

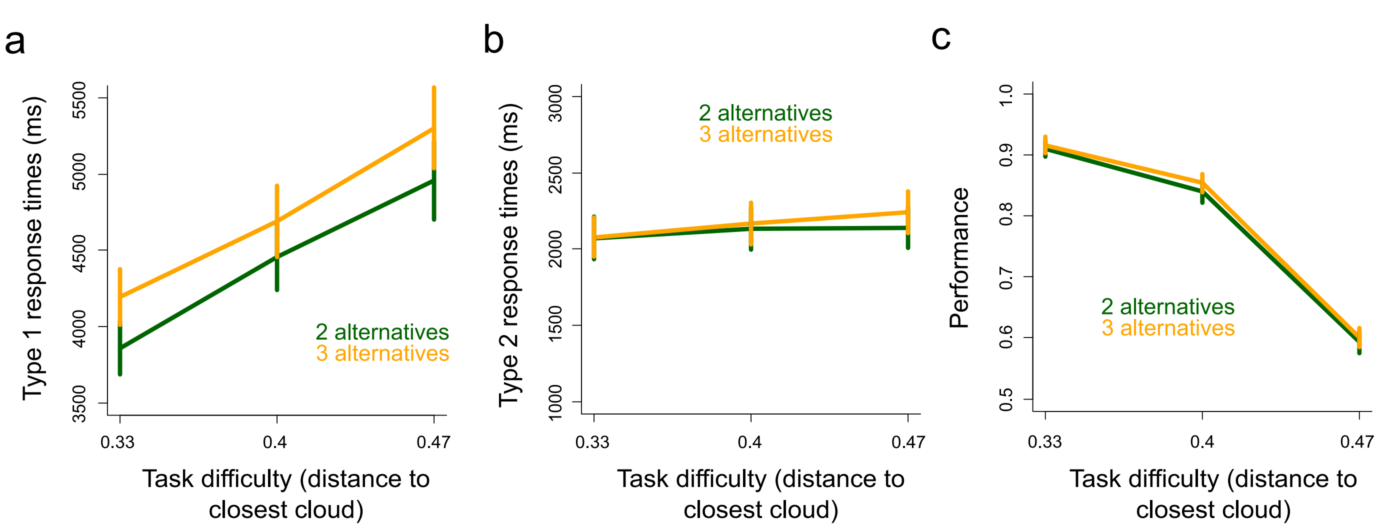

Fig. 10. (a) Both task difficulty and the presence of the dud alternative increased response times in the decisions. (b) Contrary to decisions, response times for the confidence report were not affected neither by task difficulty nor the amount of alternatives. (c) Performance decreased as the task difficulty increased, whereas the dud alternative did not affect it.

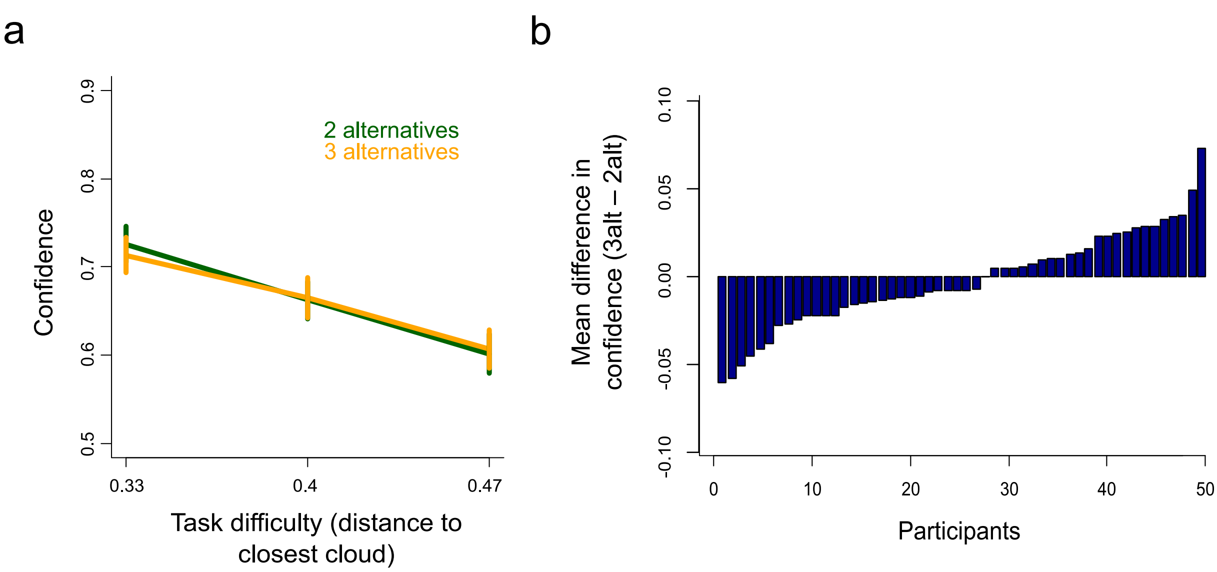

Fig. 11. Effect of number of alternatives on confidence. (a) Confidence diminished with task difficulty and is not affected by the presence of an irrelevant alternative. (b) Confidence difference between 3 and 2 alternative conditions, per participant.

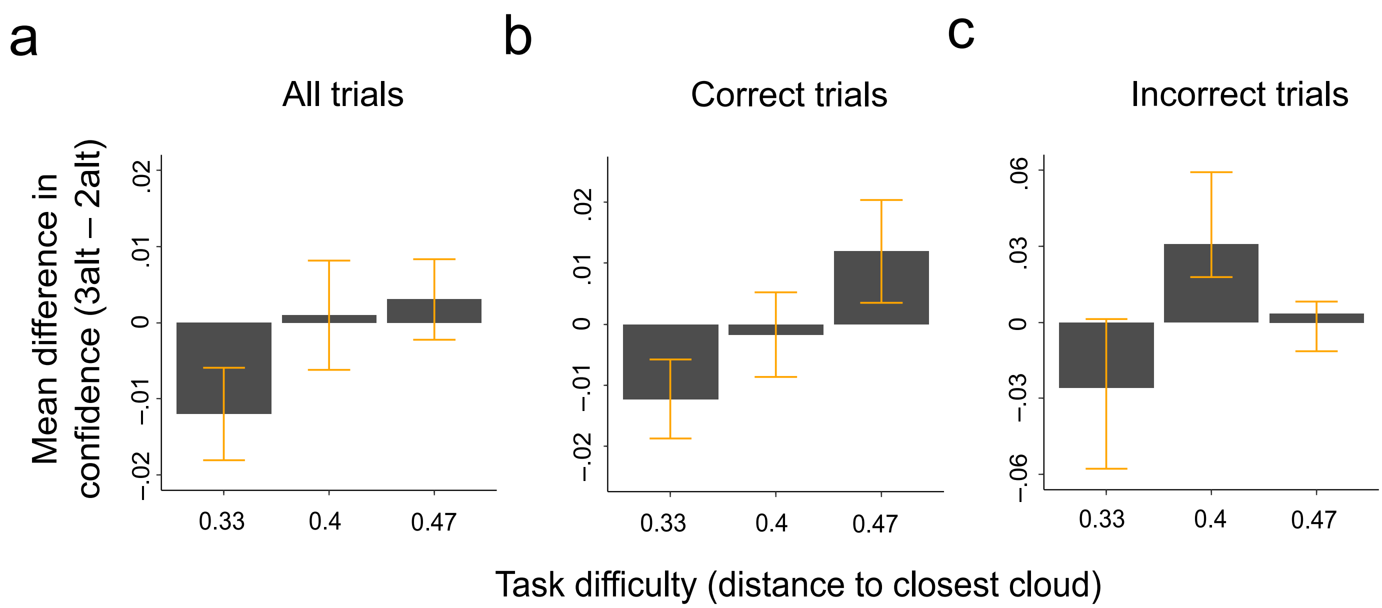

Fig. 12. The impact of the third alternative on confidence dissociated by correct and incorrect responses. (a) Considering all trials, the third alternative did not increase confidence at any level of difficulty. Moreover, the third alternative did not have an effect on confidence neither when considering (b) correct trials nor (c) incorrect trials.

The design was similar to the design of Experiment 1, with the difference that the stimuli remained on screen up to the time that the participant made the decision. To inform their choice, participants clicked directly on the geometrical figures.

6.1.2. *Data analysis*

We excluded trials with RT lower than 200 ms for both decisions and confidence reports. We also excluded trials with RT >20 s (both in decisions 1 and confidence reports) as in Experiment 3. With this filter one subject was completely excluded because of larger response times. The reported results do not change if we include this data. We conducted repeated measures ANOVAs and Tukey HSD post hoc tests (both with the Statistica software) to estimate the impact of the difficulty of the task and the dud alternatives in RT, performance and confidence. Our predefined *p*-value for statistical significance was 0.05. The ANOVA for incorrect trials was done with the highest levels of difficulty (0.9, 0.93, 0.95) because there were not enough incorrect trials in easier conditions (0.7 and 0.8).

6.2. *Results*

6.2.1. *Response times*

Task difficulty had a significant effect on decisions RT ($F_{4,\ 128} = 27.49$, $p < 0.000001$, $\eta_p^2 = 0.46$) (Fig. 13a). The inclusion of a third alternative did not affect decisions RT ($F_{1,\ 32} = 0.39$, $p = 0.54$, $\eta_p^2 = 0.02$). No interaction was found between the number of alternatives and task difficulty ($F_{4,\ 128} = 1.12$, $p = 0.35$, $\eta_p^2 = 0.03$). Task difficulty affected confidence reports RT ($F_{4,\ 128} = 5.79$, $p = 0.0003$, $\eta_p^2 = 0.15$). The number of alternatives did not impact confidence reports RT ($F_{1,\ 32} = 0.49$, $p = 0.49$, $\eta_p^2 = 0.015$). An interaction was found between task

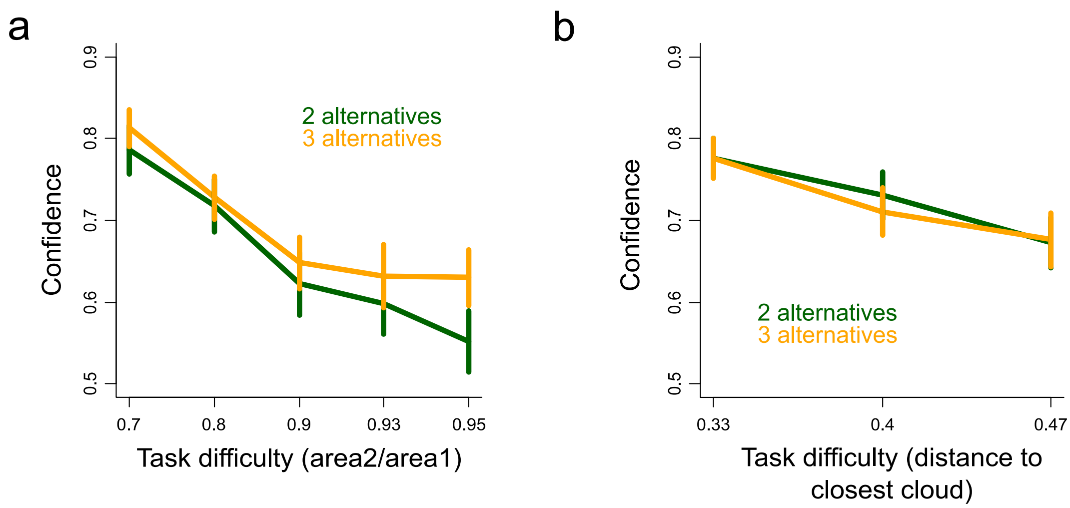

Fig. 13. Both experiments show converging evidence that the dud-alternative effect is independent of the presentation time of the stimuli. a) The dud-alternative effect was found for the size discrimination task even if stimuli remained on screen up to the participants' response. b) Despite restricting stimuli presentation time up to 1 s in the target categorization task, no dud-alternative effect was found.

difficulty and the number of alternatives regarding confidence reports RT ($F_{4,\ 128} = 2.84$, $p = 0.03$, $\eta_p^2 = 0.08$).

6.2.2. *Performance*

Performance, as expected, decreased as the task became more difficult ($F_{4,\ 128} = 92.64$, $p < 0.000001$, $\eta_p^2 = 0.74$). The presence of a third alternative did not have an effect on performance ($F_{1,\ 32} = 2.80$, $p = 0.10$, $\eta_p^2 = 0.08$). We did not find an interaction between task difficulty and the number of alternatives ($F_{4,\ 128} = 1.45$, $p = 0.22$, $\eta_p^2 = 0.04$).

We again analyzed whether the shape of the third alternative had an effect on the shape that the participant picked ("square" vs "circle" choices). Comparing the 2-alternative condition and the 3-alternative condition when the dud-alternative was a square shows no effect regarding "square" choices ($F_{1,\ 32} = 0.05$, $p = 0.82$, $\eta_p^2 = 0.002$). The presence of a third alternative that was a circle did not affect the proportion of "circle" choices either ($F_{1,\ 32} = 1.11$, $p = 0.3$, $\eta_p^2 = 0.025$).

6.2.3. *Confidence*

Confidence, as expected and in line with performance, decreased with task difficulty ($F_{4,\ 128} = 49.02$, $p < 0.000001$, $\eta_p^2 = 0.61$) (Fig. 13a). We replicated the dud-alternative effect found in Experiments 1 and 2 since confidence increased in the 3-alternative condition ($F_{1,\ 32} = 20.28$, $p = 0.00008$, $\eta_p^2 = 0.39$) (Fig. 13a). Also, there was an interaction between the number of alternatives and task difficulty ($F_{4,\ 128} = 3.03$, $p = 0.02$, $\eta_p^2 = 0.09$) (Fig. 13a). The effect was found for both correct ($F_{1,\ 32} = 16.79$, $p = 0.0003$, $\eta_p^2 = 0.34$) and incorrect ($F_{1,\ 22} = 7.87$, $p = 0.01$, $\eta_p^2 = 0.26$) trials. Comparing the confidence increment between correct and incorrect trials (on the 3 highest levels of difficulty) no differences were found ($F_{1,\ 22} = 0.88$, $p = 0.36$, $\eta_p^2 = 0.04$).

Regarding the size of the third alternative, we found a significant effect ($F_{6,\ 186} = 2.67$, $p = 0.02$, $\eta_p^2 = 0.08$). A Tukey HSD post-hoc test shows that the differences are between the 2-alternative condition and the conditions where the third alternative adopted the 0.07 ($p = 0.04$) and the 0.28 ($p = 0.008$) of the area of the stimulus 1.

As in Experiment 1, we did not find an effect of the shape of the third alternative on confidence: when considering if the dud alternative was a circle or a square only on the 3-alternative condition, no difference was found in confidence neither for "circle" choices ($F_{1,\ 21} = 0.73$, $p = 0.4$, $\eta_p^2 = 0.03$) nor for "square" choices ($F_{1,\ 27} = 0.43$, $p = 0.52$, $\eta_p^2 = 0.016$).

7. Experiment 5

As mentioned previously, to evaluate if presentation time of the stimuli is the main factor driving the dud-alternative effect we also conducted the target categorization task of Experiment 3 but with restricted presentation time: stimuli were present on screen up to 1 s as in Experiments 1 and 2.

7.1. *Materials & methods*

7.1.1. *Participants & experimental design*

31 participants took part in this experiment (83.9% females; mean age = 27.42, sd = 7.64). Participants should read and accept an informed consent prior to the realization of the experiment. All participants reported no psychiatric, illegal drug consumption or neurological history. The tasks' instructions were written on screen prior to the experiment. All participants completed the experiment on a computer.

The design was the same as Experiment 3, with the difference that the stimuli remained on screen up to 1 s.

7.1.2. *Data analysis*

We excluded trials with RT lower than 200 ms for both decisions and confidence reports. We also excluded trials with RT >10 s (both in decisions and confidence reports) as in Experiments 1 and 2. Including the discarded data does not modify the results. We conducted repeated measures ANOVAs and Tukey HSD post hoc tests (both with the Statistica software) to estimate the impact of the difficulty of the task and the dud alternatives in RT, performance and confidence. Our predefined *p*-value for statistical significance was 0.05.

7.2. *Results*

7.2.1. *Response times*

Task difficulty had a significant effect on decisions RT ($F_{2,\ 60} = 13.73$, $p = 0.00001$, $\eta_p^2 = 0.31$). The inclusion of a third alternative made decisions RT larger ($F_{1,\ 30} = 4.84$, $p = 0.04$, $\eta_p^2 = 0.14$). No interaction was found between the number of alternatives and task difficulty for decisions RT ($F_{2,\ 60} = 1.52$, $p = 0.23$, $\eta_p^2 = 0.05$). Task difficulty did not affect confidence reports RT ($F_{2,\ 60} = 1.66$, $p = 0.20$, $\eta_p^2 = 0.05$). The number of alternatives did not impact confidence reports RT ($F_{1,\ 30} = 0.67$, $p = 0.42$, $\eta_p^2 = 0.02$). No interaction was found between task difficulty and the number of alternatives regarding confidence reports RT ($F_{2,\ 60} = 0.44$, $p = 0.65$, $\eta_p^2 = 0.01$).

7.2.2. *Performance*

Performance decreased with task difficulty ($F_{2,\ 60} = 155.77$, $p <$

0.000001, $\eta_p^2 = 0.84$). The presence of a third alternative did not have an effect on performance ($F_{1, 30} = 0.035$, $p = 0.85$, $\eta_p^2 = 0.001$). We did not find an interaction between task difficulty and the number of alternatives ($F_{2, 60} = 0.52$, $p = 0.6$, $\eta_p^2 = 0.02$).

7.2.3. Confidence

Confidence decreased with increasing levels of task difficulty ($F_{2, 60} = 31.47$, $p < 0.000001$, $\eta_p^2 = 0.51$) (Fig. 13b). As in Experiment 3, confidence was not affected by the presence of a dud-alternative ($F_{1, 30} = 1.15$, $p = 0.29$, $\eta_p^2 = 0.04$) (Fig. 13b). There was not an interaction between the number of alternatives and task difficulty ($F_{2, 60} = 1.69$, $p = 0.19$, $\eta_p^2 = 0.05$) (Fig. 13b).

7.3. Discussion

Both Experiments 4 and 5 show converging evidence that the dud-alternative effect is not dependent on the presentation time of the stimuli, as the effect was replicated in Experiment 4 and, again, not found in the target categorization task of Experiment 5.

The dud alternatives did not affect performance nor RT in the size discrimination task, adding more evidence to the previous claims of Experiment 1 and 2 that the dud-alternatives did not make the task harder. Regarding RTs for the target categorization task, we found again that they were longer in the 3-alternative condition (as in Experiment 3). This could reflect that people were in fact more uncertain when the third alternative was present, thus canceling out a possible dud-alternative effect because of a decision process inherently more difficult.

Experiment 4 shows no “attraction effect”: people were equally likely to choose a square when the third alternative was a square when compared to the 2-alternative condition, and the same holds for “circle” choices. Moreover, this experiment replicates the findings of Experiment 1 in the sense that the shape of the dud-alternative did not affect confidence reports. In consequence, both experiments suggest that the dud-alternative effect is independent of the attraction effect.

8. General discussion

In the present study, we investigated whether irrelevant alternatives increase confidence in a perceptual decision. The underlying motivation was that this kind of effect was reported in likelihood judgments (Windschitl & Chambers, 2004), line-up identifications (Charman et al., 2011) and associative memory (Hanczakowski et al., 2014) but never tested in perceptual confidence contexts. Furthermore, most accepted computational models of confidence in perceptual decisions (H.-H. Li & Ma, 2020;Meyniel et al., 2015; Sanders et al., 2016) state that the presence of irrelevant alternatives should not modify it or, in any case, should decrease it. Contrary, our main finding is that these alternatives can increase confidence in perceptual decisions, a result that seems at odds with normative models of confidence.

These weak alternatives, however, do not impact RTs (Experiments 1 and 2) or performance. These null effects are important because they discard that changes in performance or overconfidence usually present in difficult tasks (Baranski & Petrusic, 1994; Gigerenzer et al., 1991) are raising confidence.

Our results from Experiment 1 and 2 cannot be explained by current views of the field proposing that confidence arises directly from the probability of being correct (Meyniel et al., 2015; Sanders et al., 2016) or direct comparisons between the more probable stimulus and the second one (H.-H. Li & Ma, 2020). This depicts why considering the effect of irrelevant alternatives (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004) in models of confidence is needed. To account for the effect in non-perceptual tasks, Charman et al. (2011), Windschitl and Chambers (2004) and Hanczakowski et al. (2014) proposed various theoretical explanations. Interestingly, the preferred one was the Contrast model, that we computationally implemented: confidence level is obtained by a series of pairwise comparisons between the chosen option and the rest. Nevertheless, our computational modeling results suggest that, although the Contrast model captures the effect of increased confidence levels when a dud-alternative is present, only the Average-residual model replicates the pattern of our data. Specifically, the Average-residual model states that people construct their confidence by first judging the support offered by the evidence for chosen option and the support offered by the average of the rest of the alternatives and then taking the difference between them. In this model having dud alternatives decreases the value of the competing options, inflating confidence.

The Contrast model faces many issues when extending its predictions to other decision contexts. First, the model predicts a constant increment in confidence when adding more than one dud-alternative (for example, a decision context with 2, 3 or more dud-alternatives), as illustrated in Supplementary Fig. 3b. This is not reasonable in extreme cases such as when an observer has to decide between a lot of alternatives (for example, adding a dud-alternative when there are already 15 dud-alternatives is not expected to raise confidence by the same amount as when there are not any dud-alternatives). The Average-residual model, conversely, predicts that increment in confidence should progressively decrease with the number of dud-alternatives, since the average of the residual alternatives will progressively change less with more dud-alternatives. Therefore, the Average-residual model is in a better position to account for the results reported in Experiment 2.

Second, when varying the size of a third alternative –ranging from a dud-alternative to a truly competitive alternative– the Contrast model still predicts that confidence should be higher in the 3-alternative condition. This is a counterintuitive prediction since it is expected that confidence will be lower in the case where 3 alternatives are equally likely to be the right decision when compared to only 2 equally likely alternatives. On the other hand, the Average-residual model predicts that confidence should be lower in that case, as expected (Supplementary Fig. 3a). This is because these models imply normalization, so a high value of a third alternative implies that the posterior probability for the chosen option also decreases, leading to a decrease in confidence. However, this decrease is not large enough to compensate for the increase in confidence by the presence of a dud-alternative in the Contrast model.

Finally, the Contrast model predicts that confidence should always increase with the presence of a dud-alternative, even if the task at hand is very easy. This is, again, a counterintuitive prediction since if a task is really easy one should reach a maximum confidence level. Therefore, the addition of a dud-alternative is not expected to have a large effect. The Average-residual model, on the other hand, predicts that the increase in confidence should be lower on easy tasks (and no increment is predicted if the two non-chosen options are equal, i.e. the both options are dud-alternatives). Indeed, according to the Average-residual mechanism, we found a null or reduced increase in confidence on easy trials on Experiment 1, 2 and 4. Taking into account all of these predictions, the Average-residual model not only is the best explanation for the dud-alternative effect, but also seems a plausible blueprint for a general mechanism for computing confidence in multi-alternative tasks in general.

Windschitl and Chambers (2004) discuss a version of the Contrast model that includes a weight in the comparison between the chosen option and the dud-alternative, thus making the increment in confidence due to this contrast flexible. With this modification, the model could improve its performance regarding the fit to our data. However, this modified Contrast model will still predict a constant, difficulty independent increment in confidence, so it is likely that the model will not be better than the Average-residual model as it is (and, consequently, will definitely not be better than an Average-residual model including a weight, to match the number of free parameters). Moreover, the modification implies that the model should have an extra free parameter for controlling the weight of the contrast, making the model more complex unnecessarily since there is already a model that provides a better fit without the need of an extra free parameter.

Prior research suggests that maintaining mental representations for all alternatives is costly (Maniscalco, Peters, & Lau, 2016). In that scenario, confidence only relies on the evidence in favor of the chosen option, also known as "positive evidence bias" (Koizumi, Maniscalco, & Lau, 2015; Maniscalco et al., 2016; Peters et al., 2017; Zylberberg, Barttfeld, & Sigman, 2012). However, unchosen options can affect confidence (H.-H. Li & Ma, 2020). Our results and model are in line with this posture, since even very weak alternatives are taken into account to make confidence judgments. Moreover, our findings also deviate from the notion that, in multi-alternative perceptual decisions, individuals making decisions only have access to the level of activity of the alternative with highest internal activity (Yeon & Rahnev, 2020). Indeed, the Average-residual model necessarily implies that the level of activity of all alternatives is accessible for the observer in order to compute confidence.

The picture, however, seems more complicated. Our results from Experiments 3 and 5 show that, in that particular task, the dud-alternative barely affects confidence. This constitutes a main challenge for the Average-residual model. Why does this experimental paradigm show no dud-alternative effect? First of all, the slower reaction times found in these experiments in the 3-alternative condition may indicate that observers were more uncertain and needed more time to make their decision; this could counteract the dud-alternative effect. Second, we tested if stimuli presentation time plays a role in the dud-alternative effect. The reason was that in perceptual tasks with brief stimuli presentations, confidence computations rely more on an internal representation (Rahnev, 2020). Thus, mechanisms acting in Experiments 1 and 2 may differ from the ones acting in Experiment 3. In fact, the dud alternative effect is more prominent in timed tasks (Windschitl & Chambers, 2004). In this line, an alternative explanation for our results might be that the dud alternative effect depends on a form of variance misperception (Zylberberg et al., 2014), an overconfidence bias found in perceptual tasks when signal is low and perceptual noise is added (M. K. Li, Lau, & Odegaard, 2018; Rahnev et al., 2011; Solovey, Graney, & Lau, 2015). The presence of dud alternatives might add perceptual noise, causing the participants to have higher levels of confidence if they do not adjust their confidence criteria accordingly. That is, if participants fixate on the third alternative, the rest of the alternatives remain at an attentional periphery, where there are effects of variance misperception and overconfidence (Winter & Peters, 2021) due to noisier percepts caused by the poorer resolution of the peripheral visual field. Larger stimuli presentations would rule out possible variance misperception effects, due to a higher signal to noise ratio. However, decreasing the stimuli duration used in Experiment 3 (Experiment 5) does not induce a raise in confidence levels in 3-alternative trials; also, the effect does not disappear with unlimited presentation time in the size discrimination task (Experiment 4). Consequently, stimuli presentation time seems not to be a factor driving the dud-alternative effect. In third place, another possible explanation might be that the dud-alternative effect in the size discrimination task is driven by the "Ebbinghaus illusion" (Roberts, Harris, & Yates, 2005): shapes are perceived larger when surrounded by small shapes. The irrelevant alternative(s) may induce such an illusion by making the chosen option appear larger to the observer as compared to when the dud-alternative is not present. Confidence therefore will be higher due to the "positive evidence bias" mentioned above: the observer has now more evidence for her choice, because her chosen option now seems larger than in the 2-alternative condition. Importantly, if this is the case, it may explain why there is not a dud-alternative effect in Experiment 3 and 5, as this illusion is not present with that stimuli. Despite appealing, the Ebbinghaus illusion explanation predicts that the increment in confidence should be more pronounced on easy trials because all of the incorrect options are small. However, we see the opposite pattern in our data: confidence level rises due to the dud-alternative presence specifically on difficult trials, consistent with the explanation proposed by the Average-residual model. Moreover, according to the Ebbinghaus illusion account, the increment in confidence should be higher with lower size values of the third alternative. However, no effect was found regarding the size of irrelevant alternatives within the 3-alternative condition: confidence increased equally irrespective of the third alternative size. This is again consistent with the Average-residual account as the model predicts a relatively constant increase in confidence as long as the alternatives are irrelevant, as in our case (Supplementary Fig. 3).

Having ruled out those possible explanations, we must note that the dud-alternative effect is already reported in a range of different tasks (Charman et al., 2011; Hanczakowski et al., 2014; Windschitl & Chambers, 2004). This implies that different mechanisms take place on the target categorization task of Experiments 3 and 5. We identify three major characteristics of this experimental paradigm that are likely causing the lack of effect. First, this task requires a probabilistic response, i.e. the target could belong to either cloud with a certain probability. This kind of probabilistic alternatives is different from the size discrimination task and also from the previous dud-alternative effect literature, where there were definite correct alternatives. This may induce a more probabilistic –and therefore, normative– use of the confidence scale, that could explain the absence of the rise in confidence levels in presence of the third cloud. Indeed, the dud-alternative effect is not present in tasks that induce concerns regarding the complementary rule (Windschitl & Chambers, 2004). Second, this task implies a direct distance judgment (between the target and the clouds) between two probable alternatives, thus making it reasonable that confidence reflects the difference between the probabilities of the best-two competing alternatives, as stated by the Diff model. Finally, there is the possibility that the third alternative was not even taken into account by the subjects for being highly irrelevant. Under this possibility, the increased response times on both target categorization experiments in the 3-alternative condition may reflect extra-time for searching for the target instead of increased uncertainty. As the distance between the third alternative and the target was not manipulated, we cannot rule out this possibility. Further research could vary this distance to evaluate whether this is a plausible explanation.

Previous research including irrelevant alternatives has shown a wide range of context effects in our decision making. The size discrimination task used here allow for the possibility of testing one particular context effect: the attraction effect (Huber et al., 1982), already reported in perceptual decision making (Trueblood et al., 2013). This effect implies that decision makers will more likely choose a specific alternative A when comparing to another alternative B when a third option simultaneously similar but inferior to A is added in the decision context. In Experiment 1, we found an attraction effect for "square" choices restricted to the highest level of difficulty (that is, participants were more likely to choose a square if the third alternative was also a square), but no clear pattern was found regarding the "circle" choices. Importantly, this effect does not interact with the increase in confidence. In Experiment 4 we did not find any attraction effect, neither in choices nor in confidence. These results suggest that the dud-alternative effect is independent of the attraction effect.

In conclusion, our results imply that —at least in some contexts— confidence in multi-alternative decisions deviates from the traditional Bayesian confidence hypothesis and from the Difference model, recently developed for decisions with multiple alternatives. Moreover, our findings suggest that unchosen options affect confidence in a way that is explained by the Average-residual model. This particular model not only gives the best explanation for the dud-alternative effect but also points out a possible mechanism for computing confidence in general multi-alternative contexts. Future computational models should consider this effect and the predictions of the Average-residual model to more accurately explain human confidence levels in perceptual decision-making.

CRediT authorship contribution statement

Nicolás A. Comay: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Gabriel Della Bella:** Software. **Pedro Lamberti:** Software. **Mariano Sigman:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Guillermo Solovey:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Project administration. **Pablo Barttfeld:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Project administration.

Declaration of Competing Interest

None.

Data availability

A link to the data and codes is on the manuscript

Acknowledgements

This research was supported by two grants from Agencia Nacional de Promoción Científica y Tecnológica (Argentina) PICT (Grant #2018-03614 and Grant #2021-0083) and a grant from Universidad de Buenos Aires (UBACyT 20020170100330BA). Nicolás A. Comay and Gabriel Della Bella have a scholarship from the National Scientific and Technical Research Council (CONICET - Argentina).

Appendix A. Supplementary data

All data and materials are stored in https://github.com/nicolascomay/confidence_dud. This includes the source code of the 5 experiments, the .jzip archives for uploading the experiments to JATOS and R code for model fitting and replication of main figures. Supplementary data to this article can be found online at [https://doi.org/10.1016/j.cognition.2023.105377].

References

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

Ariely, D., & Wallsten, T. S. (1995). Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes, 63*(3), 223–232. https://doi.org/10.1006/obhd.1995.1075

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics, 55*(4), 412–428. https://doi.org/10.3758/BF03205299

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior, 35*(6), 479–500. https://doi.org/10.1007/s10979-010-9261-1

Drugowitsch, J., Moreno-Bote, R., & Pouget, A. (2014). Relation between belief and performance in perceptual decision making. *PLoS One, 9*(5), Article e96511. https://doi.org/10.1371/journal.pone.0096511

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00443

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. https://doi.org/10.3758/BF03196546

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506–528. https://doi.org/10.1037/0033-295X.98.4.506

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews, 55*, 88–97. https://doi.org/10.1016/j.neubiorev.2015.04.006

Hanczakowski, M., Zawadzka, K., & Higham, P. A. (2014). The dud-alternative effect in memory for associations: Putting confidence into local context. *Psychonomic Bulletin & Review, 21*(2), 543–548. https://doi.org/10.3758/s13423-013-0497-x

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*(1), 90. https://doi.org/10.1086/208899

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. https://doi.org/10.1098/rstb.2012.0037

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015

Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics, 77*(4), 1295–1306. https://doi.org/10.3758/s13414-015-0843-3

Lange, K., Kühn, S., & Filevich, E. (2015). "Just another tool for online studies" (JATOS): An easy solution for setup and management of web servers supporting online studies. *PLoS One, 10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0130834

Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications, 11*(1), 2004. https://doi.org/10.1038/s41467-020-15581-6

Li, M. K., Lau, H., & Odegaard, B. (2018). An investigation of detection biases in the unattended periphery during simulated driving. *Attention, Perception, & Psychophysics, 80*(6), 1325–1332. https://doi.org/10.3758/s13414-018-1554-3

Luce, R. D. (1979). *Individual choice behavior: A theoretical analysis*. Greenwood Press.

Ma, W. J. (2019). Bayesian decision models: A primer. *Neuron, 104*(1), 164–175. https://doi.org/10.1016/j.neuron.2019.09.037

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science, 2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition, 21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics, 78*(3), 923–937. https://doi.org/10.3758/s13414-016-1059-x

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron, 88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience, 28*(17), 4435–4445. https://doi.org/10.1523/JNEUROSCI.5564-07.2008

Northoff, G., & Mushiake, H. (2020). Why context matters? Divisive normalization and canonical microcircuits in psychiatric disorders. *Neuroscience Research, 156*, 130–140. https://doi.org/10.1016/j.neures.2019.10.002

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., … Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour, 1*(7), 0139. https://doi.org/10.1038/s41562-017-0139

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864–901. https://doi.org/10.1037/a0019737

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience, 19*(3), 366–374. https://doi.org/10.1038/nn.4240

Rahnev, D. (2020). Confidence in the real world. *Trends in Cognitive Sciences, 24*(8), 590–591. https://doi.org/10.1016/j.tics.2020.05.005

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences, 41*, Article e223. https://doi.org/10.1017/S0140525X18000936

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience, 14*(12), 1513–1515. https://doi.org/10.1038/nn.2948

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review, 120*(3), 697–719. https://doi.org/10.1037/a0033152

Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception, 34*(7), 847–856. https://doi.org/10.1068/p5273

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron, 90*(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception, & Psychophysics, 77*(1), 258–271. https://doi.org/10.3758/s13414-014-0769-1

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science, 24*(6), 901–908. https://doi.org/10.1177/0956797612464241

Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(1), 198–215. https://doi.org/10.1037/0278-7393.30.1.198

Winter, C. J., & Peters, M. A. K. (2021). Variance misperception under skewed empirical noise statistics explains overconfidence in the visual periphery [preprint]. *Neuroscience.*. https://doi.org/10.1101/2021.01.08.425966

Yeon, J., & Rahnev, D. (2020). The suboptimality of perceptual decision making with multiple alternatives. *Nature Communications, 11*(1), Art. 1. https://doi.org/10.1038/s41467-020-17661-z

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience, 6*. https://doi.org/10.3389/fnint.2012.00079

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition, 27*, 246–253. https://doi.org/10.1016/j.concog.2014.05.012