Article:

Margoni, F. and Walkinshaw, N. orcid.org/0000-0003-2134-6548 (2024) Subjective logic as a complementary tool to meta-analysis to explicitly address second-order uncertainty in research findings: a case from infant studies. Infant Behavior and Development, 76. 101978. ISSN 0163-6383

https://doi.org/10.1016/j.infbeh.2024.101978

Infant Behavior and Development 76 (2024) 101978

Contents lists available at ScienceDirect

Infant Behavior and Development

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

Subjective logic as a complementary tool to meta-analysis to explicitly address second-order uncertainty in research findings: A case from infant studies

Francesco Margoni [a,b,1,*], Neil Walkinshaw [c,2]

[a] Department of Social Studies, University of Stavanger, Norway
[b] Department of Psychology, University of Oslo, Norway
[c] Department of Computer Science, The University of Sheffield, United Kingdom

ARTICLE INFO

Keywords:
Subjective logic
Meta-analysis
Uncertainty
Replication
Infancy research

ABSTRACT

Any experiment brings about results and conclusions that necessarily have a component of uncertainty. Many factors influence the degree of this uncertainty, yet they can be overlooked when drawing conclusions from a body of research. Here, we showcase how *subjective logic* could be employed as a complementary tool to meta-analysis to incorporate the chosen sources of uncertainty into the answer that researchers seek to provide to their research question. We illustrate this approach by focusing on a body of research already meta-analyzed, whose overall aim was to assess if human infants prefer prosocial agents over antisocial agents. We show how each finding can be encoded as a subjective opinion, and how findings can be aggregated to produce an answer that *explicitly* incorporates uncertainty. We argue that a core feature and strength of this approach is its *transparency* in the process of factoring in uncertainty and reasoning about research findings. Subjective logic promises to be a powerful complementary tool to incorporate uncertainty explicitly and transparently in the evaluation of research.

1. Introduction

A central and often unavoidable component of any set of empirical findings is *uncertainty* (Camerer et al., 2018; Makel, Plucker, & Hegarty, 2012; Open Science Collaboration, 2015; cf. also Breznau et al., 2022). Generally, researchers do not have direct access to a 'ground truth', and instead are forced toward best estimates or ways to weigh evidence in favor of or contrary to a given hypothesis or theory. This situation can be especially challenging when a theoretical claim must be evaluated based on multiple experiments, each bringing about its own uncertainty.

When discussing probabilistic (and therefore inherently uncertain) phenomena, it can be helpful to discuss the underlying

* Correspondence to: University of Stavanger, Institutt for Sosialfag, 4021 Stavanger, Norway.
E-mail address: francesco.margoni@uis.no (F. Margoni).
[1] ORCID ID 0000-0002-8109-228X
[2] ORCID ID 0000-0003-2134-6548

https://doi.org/10.1016/j.infbeh.2024.101978
Received 14 November 2023; Received in revised form 8 July 2024; Accepted 10 July 2024
Available online 31 July 2024
0163-6383/

uncertainty in terms of two dimensions (Jøsang, 2001, 2016; cf. also Shafer, 1976). The 'first-order' uncertainty pertains to the intrinsic variability or randomness of the phenomenon itself (when measured) – e.g., the probability that a hypothesis is confirmed by an experiment. The 'second-order' uncertainty pertains to our confidence in that probability (Der Kiureghian & Ditlevsen, 2009). The experiment in question might be a tightly-controlled replication that has been conducted on thousands of participants in the past, or it might be a loosely controlled experiment conducted for the first time on a small handful of participants; second-order uncertainty pertains to the different levels of confidence with which we would interpret or report the findings of the two experiments.

In practice, second-order uncertainty can be related to the use of different methodologies, measures, and procedures across studies, as well as to numerous factors that are often simply regarded as part of the natural and unavoidable heterogeneity between studies, such as the different degree of expertise and training of experimenters working in different laboratories. All these factors contribute to make the interpretation of the findings less certain, and the conclusions less firm. When faced with entire bodies of research, spanning decades, it is at least as important to factor-in the inherent second-order uncertainty as it is to analyze and aggregate the primary outcomes.

Initiatives have been taken to reduce uncertainty and enhance the overall trustworthiness of the findings and the possibility of meaningfully pooling together the evidence at disposal. These initiatives have been taken across many fields (Cumming, 2014; Goodman, Fanelli, & Ioannidis, 2016; Nosek et al., 2015; Szucs & Ioannidis, 2017; Wilson, Harris, & Wixted, 2020). Many journals now require reporting of additional metrics to help better evaluate the findings and their accuracy (e.g., confidence intervals, statistical power of the tests employed), and Open Data initiatives work to enhance the transparency of the reporting (Asendorpf et al., 2013; Chambers, 2013; Nosek & Lakens, 2014; see also Grand, Rogelberg, Banks, Landis, & Tonidandel, 2018; van't Veer & Giner-Sorolla, 2016; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Meta-analysis and Bayesian techniques can also be employed to reason more thoroughly about uncertainty (Morey & Rouder, 2011; Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017; Tsuji, Bergmann, & Cristia, 2014). Lastly, large-scale multi-lab replication studies are increasing in number as a possible answer to the request for greater certainty (Frank et al., 2017; Uhlmann et al., 2019).

Here, we introduce a tool that is *complementary* to meta-analysis, providing researchers with the chance to reason explicitly about second-order uncertainty in research findings. Our aim is not to offer a new technique to reduce uncertainty. We acknowledge that research findings necessarily have a (potentially large) component of (first-order) uncertainty, arising for instance from the experimental setting. We propose to use *subjective logic* (Jøsang, 2016) to enable researchers to explicitly factor-in different sources of uncertainty, both for the interpretation of single findings and in the interpretation of a larger body of research. This approach provides researchers with a more accurate, explicit and transparent assessment of the evidence *in light of* (and not *despite of*) the uncertainty underlying the findings.

In what follows, we briefly introduce subjective logic, define core concepts of uncertainty, subjective opinion, and projected probability, and show how a subjective opinion can be built and graphically displayed. Also, we show how subjective opinions can be combined by using belief fusion operators. Next, we show subjective logic in practice, by applying it to the analysis of a set of findings already meta-analyzed (cf. Margoni & Surian, 2018) whose overall aim was to provide an answer to the question of whether children behave consistently with a preference for prosociality already in the first two years of life (which would ostensibly indicate the presence of an early-emerging precursor of moral sense in humans).

This application seeks to offer an example for researchers to draw on if they aim to use subjective logic as a complementary tool to meta-analysis to reason about uncertainty. We illustrate the process step by step, starting with examples of parameters definition and uncertainty encoding. Next, we show how subjective opinions, the very building blocks of subjective logic, can be fused together to provide an answer to the main question (in our application, the question of whether infants possess a moral sense, that is, a preference for approaching prosocial agents over antisocial agents). We also showcase how opinions can be fused together in different groups separated along the line of some variables of interest, and how fused opinions can then be compared to each other. Last, we illustrate how to assess whether the subjective opinion expressing the belief in an early-emerging moral sense changed over time since the first original study conducted in this field. We conclude by assessing and clarifying the main strengths and purposes of this tool, as well as its main intrinsic limitations. As a final remark, we again stress that subjective logic should not be intended as a substituting tool but as a complementary tool to meta-analysis, where its main strength is the possibility offered to researchers of focusing and reasoning more carefully and explicitly about uncertainty in research findings, a possibility somehow hindered by the most classical approaches to statistical testing, especially when the sources of uncertainty are numerous.

2. Modelling subjective opinions

Subjective logic is a type of probabilistic logic where probability values can be expressed with degrees of uncertainty, and it has been developed as a framework for modelling and analyzing situations characterized by uncertainty or incomplete knowledge (Jøsang, 2016). Subjective logic allows to express belief or disbelief about a certain proposition and the degree of trust in the available evidence. This framework has been already employed to handle problems in different fields, such as law (Jøsang and Bondi, 2000), determining trustworthiness within multi-agent systems (Cerutti, Kaplan, Norman, Oren, & Toniolo, 2015; Oren, Norman, & Preece, 2007), vehicular communication networks (Dietzel, van der Heijden, Decke, & Kargl, 2014), and empirical software engineering (Walkinshaw

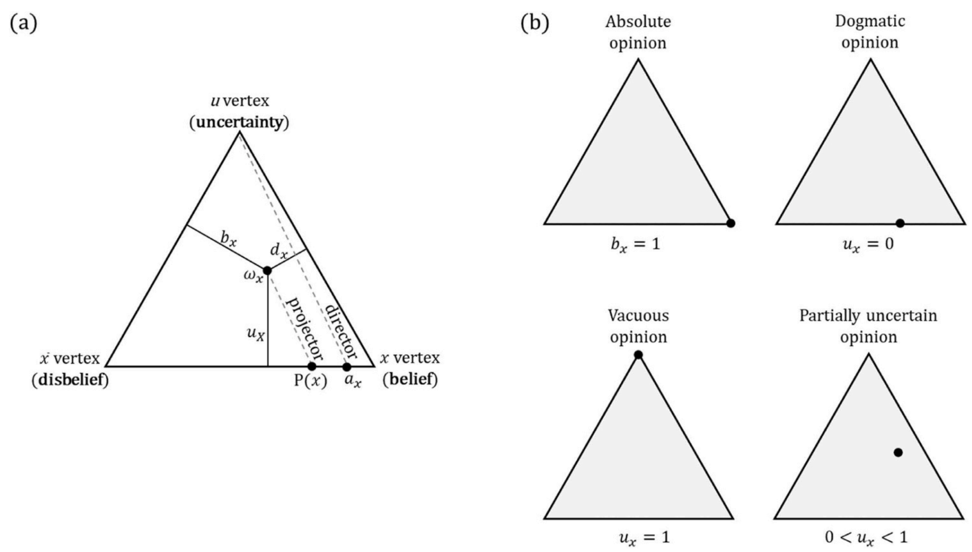

Fig. 1. *Barycentric Triangle Visualization of a Subjective Opinion. Note.* Panel (a) displays the basic dimensions of a barycentric triangle, where the u vertex represents the maximum for uncertainty, and the other two vertices represent the maximum for disbelief and belief respectively. The projected probability from a given opinion ω_x is computed from the belief, disbelief and uncertainty masses, as well as the prior probability of x (see Definition 2, Section 2.2). Panel (b) shows how an absolute (i.e. boolean true or false), a dogmatic (i.e. traditional probability), a vacuous (no belief or disbelief, just uncertainty), and a partially uncertain opinion respectively can be graphically visualized (cf. Jøsang, 2016).

& Hierons, 2023; Walkinshaw & Shepperd, 2020).

Building on Dempster Shafer theory (Dempster, 1968; Shafer, 1976), Jøsang (2016) offers a formalism to model and reason about *beliefs*, where beliefs can be understood as propositions about a certain phenomenon (for instance, the belief that infants possess vs. do not possess a moral sense) and characterized as *probabilities*. What is novel to this approach is that these probabilities can be associated with explicit measures of *uncertainty*, where uncertainty refers to lack of relevant information or knowledge that reduces one's ability to reach a firm conclusion about the quantification of the probability or belief. The resulting outcome of this modelling is called *subjective opinion*. Here, we focus on binomial opinions (boolean true/false),[3] though multinomial or hyper-opinions can also be modelled (see Jøsang, 2016). As we show in our example, the answer to the question about whether infants possess a moral sense can be represented as a binomial (true/false) subjective opinion.

2.1. *Definition 1: binomial subjective opinions*

A binomial subjective opinion refers to the truth of some x, where $\mathbb{X} = \{x, \overline{x}\}$ is a binary domain with binomial random variable $X \in \mathbb{X}$. A binomial subjective opinion can be modelled as the ordered quadruplet $\omega_x = (b_x, d_x, u_x, a_x)$. The parameters are defined as follows:

b_x: *belief mass* in support of x being true, i.e. $X = x$.

d_x: *disbelief mass* in support of x being false, i.e. $X = \overline{x}$.

u_x: *uncertainty mass* representing the 'vacuity of evidence'.

a_x: *base rate* or *prior probability* of x without having observed any evidence. An 'uninformed prior' would amount to $a_x = 0.50$ (i.e. without any additional information, x and $\overline{x}$ are equally likely).

There is a requirement that $b_x + d_x + u_x = 1$. This formalization leads to a few special classes of binomial opinions:

$b_x = 1$ or $d_x = 1$: *absolute opinion*, equivalent to Boolean true or false.

$u_x = 0$: *dogmatic opinion*, which also corresponds to a traditional probability.

$u_x = 1$: *vacuous opinion*, with zero belief or disbelief mass.

$0 < u_x < 1$: *partially uncertain opinion*.

It is worth noting that conventional probabilistic statements all fall within the second category. The statement "50% of infants reached for the prosocial agent" does not in and of itself imply any *second-order* uncertainty (though it naturally implies a significant level of first-order uncertainty). It is commonly left to the reader to factor in their own second-order uncertainty (u_x) by reporting ancillary data such as the number of participants and statistical test results, measures of the overall quality of the experiments, etc.

[3] We acknowledge that this focus represents quite an important restriction, but it was chosen to make our argument and presentation of the tool easy to follow. However, a focus on binomial opinions can be made flexible by adopting a smallest effect size of interest approach (Lakens, 2017), where true would be anything of interest for a given community of researchers (thus using specific null hypotheses instead of a nil hypothesis; Meehl, 1967).

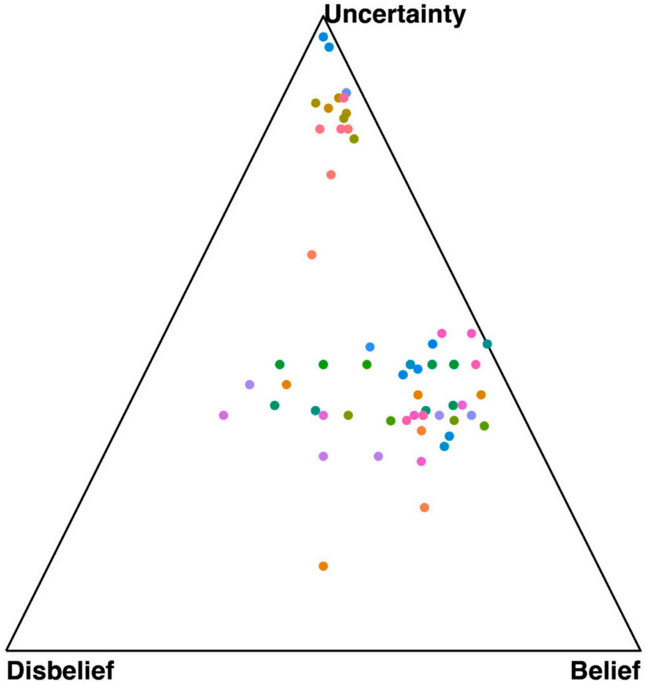

Fig. 2. *Barycentric Triangle Visualization of the Subjective Opinions for the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis***.** *Note*. The color legend is provided in Fig. 3.

The requirement that $b_x + d_x + u_x = 1$ means that it is possible to represent a binomial opinion through only two of the three values, where the third value is implicit. For the sake of clarity, we use a complete formulation.

Binomial opinions can be visualized as points in a barycentric coordinate system of three axes (belief, disbelief, and uncertainty), graphically represented by an equilateral triangle where each of the vertices represents absolute belief, disbelief and uncertainty respectively (see Fig. 1).

2.2. *Definition 2: projected probability*

Given some opinion ω_x we can define the *projected probability* $\mathrm{P}(x)$ as equal to $b_x + a_x u_x$ (see panel a in Fig. 1). To give an example, if $\omega_x = (0.40,\ 0.20,\ 0.40,\ 0.90)$ then $\mathrm{P}(x) = 0.76$. Subjective binomial opinions can be mapped and visualized as *beta-distributions*. These are the continuous version of the binomial distribution, thus appropriate for modelling probabilities.

2.3. *Definition 3: beta distributions*

The Beta distribution is defined in terms of parameters α and β. The probability density function of the distribution can be expressed as follows:

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$$

where $\alpha > 0, \beta > 0$ and $B(\alpha,\beta)$ is the beta function.

Subjective binomial opinions can be mapped to parameters α and β (see Jøsang, 2016), such that they can be also interpreted as a continuous distribution where the probability is represented and can be visualized as density. Given a subjective opinion ω_x we can define $\alpha = \frac{b_x{}^*W}{u_x} + a_x{}^*W$ and $\beta = \frac{d_x{}^*W}{u_x} + (1 - a_x)^*W$. Here, W represents the 'non-informative prior weight'. Jøsang (2016) sets this to $W = 2$, which ensures that the Beta density function is uniform when the prior probability is non-informative (i.e., $a_x = 0.5$). Now, when opinions are mapped and represented in that way (see Fig. 2 panel b; for an opinion visualization demo, see https://folk.universitetetioslo.no/josang/sl/BV.html), one can observe uncertainty in the flatness of the distribution, whereas well-defined peaks indicate greater levels of certainty about the given probabilities.

Last, different opinions can be fused together by using several belief fusion operators. Here, we focus on the *Weighted Belief Fusion* (WBF) operator. Our investigation of infant social preferences is fusing together multiple opinions representing multiple research findings, and the WBF operator gives specific weights to each opinion as a function of its relative confidence or certainty.

2.4. Definition 4: Weighted Belief Fusion (WBF)

Let $\omega_X^A = (b_X^A, d_X^A, u_X^A, a_X^A)$ denote the opinion of $A \in \mathbb{A}$ over X where $\mathbb{A}$ is a finite set of agents. The WBF of these opinions $\omega_X^{\hat{\diamond}\mathbb{A}} = (b_X^{\hat{\diamond}\mathbb{A}}, d_X^{\hat{\diamond}\mathbb{A}}, u_X^{\hat{\diamond}\mathbb{A}}, a_X^{\hat{\diamond}\mathbb{A}})$ is defined as[4]:

$$b_X^{\hat{\diamond}\mathbb{A}} = \frac{\Sigma_{A\in\mathbb{A}} b_X^A(x)(1-u_X^A)\prod_{A'\in\mathbb{A}, A'\neq A} u_X^{A'}}{(\Sigma_{A\in\mathbb{A}}\prod_{A'\neq A} u_X^A) - |\mathbb{A}| \bullet \prod_{A\in\mathbb{A}} u_X^A}$$

$$u_X^{\hat{\diamond}\mathbb{A}} = \frac{(|\mathbb{A}| - \Sigma_{A\in\mathbb{A}} u_X^A) \bullet \prod_{A\in\mathbb{A}} u_X^A}{(\Sigma_{A\in\mathbb{A}}\prod_{A'\neq A} u_X^A) - |\mathbb{A}| \bullet \prod_{A\in\mathbb{A}} u_X^A}$$

$$d_X^{\hat{\diamond}\mathbb{A}} = 1 - (b_X^{\hat{\diamond}\mathbb{A}} + u_X^{\hat{\diamond}\mathbb{A}})$$

$$a_X^{\hat{\diamond}\mathbb{A}} = \frac{\Sigma_{A\in\mathbb{A}} a_X^A(x)(1-u_X^A)}{|\mathbb{A}| - \Sigma_{A\in\mathbb{A}} u_X^A}$$

3. Subjective logic in practice: an example from infancy research

Here, we illustrate an application of the subjective logic tool to *complement* meta-analysis by applying it to the study of human infant social preferences. Our application seeks to be illustrative without being necessarily informative to the field of infancy research. We showcase how subjective logic can be used to accommodate different sources of uncertainty found in the body of research meta-analyzed by Margoni and Surian (2018). We illustrate how each finding can be modelled as a subjective opinion, and how the resulting opinions can be fused together to provide an answer to the core research question of our example ('do infants have a preference for prosociality?'). This answer is not only expressed as a proportion (i.e. n% of infants preferring a prosocial agent to an antisocial one) but it is also a clearly visualizable result of the aggregation of the different sources of uncertainty modelled in the subjective opinions. Last, we show how this tool can be employed to model the cumulative nature of research, and show if and how the answer to the research question, expressed as a subjective opinion, changed over time, with accumulating evidence and (un)certainty.

3.1. The problem

Spearheaded by the initial work of Hamlin, Wynn, and Bloom (2007), one of the most influential bodies of work on infant cognition is seeking evidence of a preference in infants and toddlers for morally good agents (e.g., helpers, fair distributors) over morally bad agents (e.g., hinderers, unfair distributors), where such evidence would be generally taken as proof of an evolved early-emerging core precursor of adult moral sense. However, a few reports claiming that the effect of an early-emerging preference for prosociality does not replicate had been appearing (e.g., Salvadori et al., 2015; see also Schlingloff, Csibra, & Tatone, 2020), which also motivated the conduction of a meta-analysis (Margoni & Surian, 2018) and, more recently, the planning of a large-scale multi-laboratory replication study (Lucca et al., 2021; see also Margoni & Shepperd, 2020).

We take advantage of this literature and the meta-analysis of Margoni and Surian (2018) to illustrate common problems related to reasoning about uncertainty and how these problems can be addressed using subjective logic. The meta-analysis estimated that approximately two out of three infants and toddlers aged 4 to 32 months, given the choice, prefer prosocial to antisocial agents. Both published and unpublished findings were retrieved, for a total of $k = 61$ effect sizes (proportion of children preferring the prosocial to the antisocial agent in each sample).

In a typical task, the child is first presented with two agents (e.g., two puppets displayed acting in front of the child) who engage in opposite behaviors toward a third agent. For example, whereas one puppet helps the third one to retrieve an object from a box, the other puppet hinders it by slamming its box shut. Next, the child is presented with both puppets (the helper and the hinderer) and encouraged to pick up one. The child's choice is taken as reflecting their social preference. However, several factors have been varied across studies. The main ones were treated as possible moderators of the main effect in the meta-analysis: (a) sample size, (b) whether the finding was published, (c) whether the study was conducted in the lab where the effect was originally found and reported (i.e. Dr. Kiley Hamlin's lab), (d) the age of the children included in the sample, (e) type of scenario illustrating prosociality vs. antisociality (helping vs. hindering, fair vs. unfair distribution, giving vs. taking or stealing), (f) type of stimuli (cartoon vs. real actors), (g) type of dependent measure (reach vs. help, if children were asked to reach for or pick up one of the two agents or if they were encouraged to

[4] Here we restrict ourselves to a situation where $(\forall A \in \mathbb{A} : u_X^A \neq 0)^{(}\exists A \in \mathbb{A} : u_X^A \neq 1)$ – other cases are defined in Van Der Heijden, Kopp, and Kargl (2018).

selectively help one of the two agents) (for details, see Margoni and Surian, 2018[5]).

To answer the core question as to whether children behave consistently with prosocial preferences early in life, we can start from the meta-analytic estimate that P = 0.68 (with 95% CI between 0.64 and 0.72) of infants and toddlers prefer prosocial agents over antisocial ones. In this paper we address the follow-on question: How confident should we be in this result? We know that there is a significant degree of heterogeneity amongst the studies. Some of this heterogeneity leads to statistical variance amongst the studies (first-order uncertainty). There were also differences in study design, some were replicated whereas others were not, and all were subject to varying degrees of scrutiny by the wider scientific community. All of these factors have a direct bearing on our confidence in their results (second-order uncertainty).

Tools such as meta-regression or Bayesian causal models can be used to reason about the statistical variances and weights that can be attributed to studies. However, they do not distinguish between first- and second-order uncertainty – cf. the question with respect to the Margoni and Surian (2018)'s analysis of whether the CI is due to heterogeneity amongst the studies, or due to other more nuanced factors that might affect our confidence in the results. If the analyst has knowledge about extrinsic factors pertaining to the experiments that could have an impact on our confidence in the results that may not be directly represented in the data, this can be difficult to encode in traditional analyses, which ultimately also fail to discriminate between these different forms of uncertainty. We return to the issue of how to generate or access knowledge about extrinsic factors in the discussion.

Ultimately, the task of aggregating and interpreting a set of results 'in the round' necessarily involves some degree of subjectivity. However, subjective logic offers a way to express and incorporate various sources of uncertainty in a manner that is both precise and explicit, in a way that can complement traditional meta-analysis techniques. The aim is not to offer a perfect non-questionable assessment of uncertainty, but to make the process of capturing and fusing together uncertainties more explicit and transparent. Instead of simply providing a superficial statistic, the aim is to provide a measure of uncertainty (or conversely confidence) that can be readily deconstructed, showing precisely how various sources of uncertainty (both first-order and second-order) have fed into it. In this paper we restrict our scope to reasoning about those sources of uncertainty that are intrinsic to the various experimental designs that are used to investigate social preferences in infancy such as the number of participants involved. However, there are in principle further potential interesting factors. The basic premises and assumptions that underpin an experimental design (for example, the notion that an infant's action indeed reflects a "moral" preference) could themselves be subject to some uncertainty. This, too, is a form of second-order uncertainty. We leave the task of capturing and quantifying this uncertainty to future work and focus here on illustrating the tool with a simple, easy to follow example for our readers.

In the next section, we show how uncertainty can be expressed precisely and explicitly by modelling findings as subjective opinions. These opinions capture all the different sources of variability related to the potential moderating factors. Crucially, the way in which different sources of uncertainty affect the overarching result is expressed both explicitly and transparently, as a primary artefact to be openly shared and discussed among researchers and laboratories, each of which can argue in favor of a different modelling of opinions. We show how the description or source code used to model subjective beliefs can be made transparent to be openly shared among researchers.

3.2. Defining parameters to calculate and encode uncertainty

Using the subjective logic approach, the tendency of infants to show a preference for prosociality is represented as a binomial subjective opinion (see Definition 1, Section 2.1). This high-level subjective opinion is obtained by fusing together a number of low-level subjective opinions, i.e. the information sources from which the answer is ultimately derived. In this context, the 'belief' and 'disbelief' values (b_x and d_x) correspond respectively to the belief or disbelief in the fact that infants prefer prosociality. The uncertainty (u_x) corresponds to the degree of confidence in this assessment.

A final component of the subjective opinion is the base-rate or prior probability (a_x). And here there are two options: We can opt to ignore the results from previous studies, and choose an 'uninformative' prior of 0.50, or we can derive our prior from previously published studies. For our approach we adopt a meta-analysis approach, whereby the individual data-points from all previous studies are aggregated to form a prior (i.e. the proportion of the cumulative number of infants with a preference for prosociality calculated with respect to the total number of infants studied so far, in the previous years). This choice is however flexible, and different experimental contexts may favour alternative approaches.

A two-phase process can be followed to populate each low-level subjective opinion. The first phase is deriving a measure of uncertainty u_x about the source of information. If, as in our example, the source of information is a statistic (i.e. a proportion expressing an effect size), we can use the confidence interval to encode uncertainty. In that case, the confidence interval must be mapped to a value between 0 and 1, where 1 means that the interval is too large for the statistic to be meaningful. Of course, as we shall see, we can add many other factors to encode and fully capture uncertainty. The crucial constraint when modelling subjective opinions is that the requirement $b_x + d_x + u_x = 1$ should be satisfied. Therefore, defining u_x will leave us with a remaining belief mass of $(1 - u_x)$, which can be split between b_x (belief mass) and d_x (disbelief mass). The second phase is then assigning values of belief and disbelief. The value e representing the source of information (e.g., an effect size statistic) should be scaled such that it is contained in the 0–1 interval. In that way, it can be mapped to b_x and d_x where $b_x = e^*(1 - u)$ and $d_x = (1 - e)^*(1 - u)$. This means that when $e = 1$ the belief mass is at its maximum and the disbelief mass at its minimum, and vice versa when $e = 0$.

5 Another interesting moderator could have been the gender of the infant, however this factor was only analyzed in a separate and subsequent meta-analysis (see Margoni, Block, Hamlin, Zmyj, & Schmader, 2023).

Before illustrating these processes in the context of our running example, a disclaimer is in order. Because our work is primarily intended as an example on how to apply subjective logic to the reasoning about uncertainty in research findings, what follows should not necessarily be taken as a theoretical contribution to the field of infant studies. That is, our main goal is to contribute to the methodological debate, by illustrating how subjective logic can be used to complement meta-analysis to reason about uncertainty, and much of the conclusions we draw on the actual data are open to discussion (a point we shall return to).

To model each of the $k = 61$ findings (retrieved and analyzed by Margoni & Surian, 2018) as subjective opinions, we started by deriving uncertainty scores.[6] For each of these scores, two end points are necessary, one representing the very best possible situation, leading to uncertainty = 0, and the other representing the worst situation, leading to uncertainty = 1. Here, as sources of uncertainty we considered the following factors:

- *95% CI* for the effect size (proportion of children preferring the prosocial agent).
- *sample size*.
- *published* (establishing whether or not the paper has been subject to peer review).
- being replicated *internally* (as opposed to externally).

The question is why these factors could help model uncertainty, and how to derive an uncertainty value for each finding based on these four factors above. We developed a formula to calculate uncertainty such that findings are more trustworthy when (a) CI is narrow (and sample size is large), (b) the study reporting the effect has been recognized by the academic community and subject to peer review, and (c) the finding has been reported by a laboratory independent from the Hamlin's lab where the effect was originally discovered. Each of these measures had to be mapped to a score between 0 and 1.

With respect to the CI factor, we can observe that findings will be at their maximum trustworthiness if CI is zero. However, we wanted to avoid the situation where a study with a very small sample size obtains a low uncertainty score because it happens to produce a narrow CI. With this in mind we also took sample size into account by setting the number of participants required for an adequately populated study to 63 (this is therefore the maximum possible score to be used to scale any number of participants to a 0–1 interval). This parameter, as well as the others, can be adjusted according to what is most reasonable within any specific domain or here, with respect to the researchers own "subjective" reasoning. We decided to set the ideal sample size at $n = 63$ because 63 is the number of participants needed to detect with a two-tailed binomial test an effect size of $P = 0.68$ (which is the estimate from the RE model in the meta-analysis) with statistical power set at 0.80 (which is what researchers in psychology customarily aspire to, see e.g., Blake & Gangestad, 2020). In sum, findings receive a low uncertainty score when N is large and CI is narrow.

3.3. *Definition 5: calculating the confidence in a statistical result (StatsScore)*

$$StatsScore = \max\left(\left(1 - \min\left(\left(\frac{n}{63}\right), 1\right)\right), (CIupper - CIlower)\right)$$

This option is far from the only reasonable option to encode sample size as a source of error. For example, a valid alternative would have been using the 'Top10' strategy (Stanley, Jarrell, & Doucouliagos, 2010) to select the most precise (or powered) 10% of the effects meta-analyzed by Margoni and Surian (2018), and based on that selection estimating an effect size of $P = 0.61$, for then subsequently calculating the ideal sample size to be $n = 162$ (that is, the minimum number of participants to detect the estimated effect size with statistical power set at 0.80 and a two-tailed binomial test). Other somewhat more arbitrary solutions could also be contemplated. For instance, it could be decided that anything below $P = 0.75$ is uninteresting from a theoretical or a practical and applied point of view. In that case, the ideal number of participants will be set accordingly. The specific threshold can be a decision of an individual researcher or a laboratory, but in principle could also be a decision based on a common agreement among researchers working in the field (an online survey could be employed to assess the degree of agreement). Next, we modelled the relative certainty associated with published (as opposed to unpublished) findings.[7]

3.4. *Definition 6: factoring in whether or not research has been peer-reviewed*

We consider a paper that has been subject to peer review and accepted for publication to be more trustworthy than a paper that has not. We indicate this with the following simple score:

$$\text{Published} = \begin{cases} 1 & \text{If the paper has not been published} \\ 0 & \text{If the paper has been published} \end{cases}$$

We thus inserted a penalty for those findings that were not peer reviewed. In our example, we chose this penalty, but it is certainly possible to develop other measures of trustworthiness. One possibility we initially considered but ended up discarding was to measure the relative impact of the findings by calculating the $\log_{500}$ of the total number of citations to the study within which the findings were reported:

[6] The source code can be found on Open Science Framework (OSF, 2023): https://osf.io/w9f8a/.

[7] Publication bias could also be an important factor to be considered. However, we reasoned that there might be many reasons for why a set of findings do not get published, only one of which is that journals resist their acceptance. Data could remain unpublished also because researchers who had collected them do not judge them reliable enough.

$$CitationScore = 1 - \min(\log_{500}(citations), 1)$$

The base of 500 would denote what we might consider to be an ideal number of citations (in the context of this literature), so any study with 500 or more citations would produce a score of 1 (which would also be the limit). Calculating the log can be used instead of the raw citation count to avoid situations where a large citation count (such as the original one by Hamlin et al., 2007 which received more than 1500 citations) could disproportionately mask differences in overall score with other unpublished studies, which might have been good in other respects but were seldom cited. However, as anticipated, we ended up with a simpler publication penalty because two major flaws of the citation score were (a) the downweighing of more recent research (which, ceteris paribus, is cited less than other research) and (b) the fact that there may be a number of reasons other than poor quality why a paper is cited less than others.

3.5. Definition 7: calculating the extent by which to moderate certainty associated with results from the hamlin lab studies (LabPenalty)

Last, a “lab penalty” was added for those studies conducted in the Hamlin’s lab, because Hamlin’s lab was where the effect was originally reported, and it has been reported now many times that internal replications succeed more often than external ones (an effect that was termed *same team science effect*, see Ioannidis, 2012). Thus, lower uncertainty was associated to the finding if it had emerged from a lab different from the one where the effect was originally discovered. Hamlin lab studies produced on average an 8% stronger effect size than non-Hamlin lab studies. In turn, this bias can be used to parameterize the penalty for internal replications (here, all the findings reported in studies conducted by Hamlin).

$$LabPenalty = \begin{cases} 0, & \textit{not from Hamlin lab} \\ \dfrac{\sum_{s \in Hamlin} xi_s}{\sum_{s \in Hamlin} ni_s} - \dfrac{\sum_{s \notin Hamlin} xi_s}{\sum_{s \notin Hamlin} ni_s}, & \textit{from Hamlin lab} \end{cases}$$

Here, the concept of ‘same team’ (or ‘research allegiance’, see Munder, Brutsch, Leonhart, Gerger, & Barth, 2013) can be operationalized in different ways. We opted for penalizing findings reported by the first author of the study in which the original effect was first reported (Hamlin et al., 2007). A different possibility would be to penalize not only the findings reported by the PI who originally claimed to have discovered the effect, but also those that were reported by the (former) students of the PI, independently of her (see for e.g., Davis et al., 2021). However, in our dataset, we did not have instances of this second type.

3.6. Definition 8: calculating u from StatsScore, Published, and LabPenalty

Having defined and derived the selected sources of uncertainty, we can now calculate a combined uncertainty score. We do this calculation in two steps. First, we calculate a score (“paper score”) that captures the uncertainties pertaining to the findings of the paper (i.e. the strength of the results, and whether they have been peer reviewed). We then apply the lab penalty to whatever this score is. *PaperScore* is simply calculated as a median value:

$$PaperScore = \text{median}\ (Published, StatsScore)$$

The final uncertainty score is computed by adding the lab penalty, and then ensuring that it does not go above 1:

$$u = \min((PaperScore + LabPenalty), 1)$$

For instance, if we take Hamlin et al. (2007) finding that 14 out of 16 infants prefer the prosocial to the antisocial agent, we can calculate the uncertainty score as follows:

$$StatsScore = \max\left(\left(1 - \left(\min\left(\frac{16}{63}, 1\right)\right)\right), (1.04 - 0.71)\right) = \max(0.75, 0.32) = 0.75$$

The paper has been subject to peer review, so there is no uncertainty arising from lack of peer review:

$$Published = 0$$

$$PaperScore = median(0, 0.75) = 0.37$$

So $u = \min(0.37 + 0.08, 1) = \min(0.45, 1) = 0.45$.

This calculation leaves us with a belief mass of 0.55, which must be divided further in belief and disbelief. In the i[th] study, the belief that infants or toddlers prefer prosocial agents to antisocial agents (i.e. that they possess a moral sense) can be computed by xi/ni where xi is the number of children preferring the prosocial agent and ni is the total sample size. This proportion must be scaled to the belief mass remaining after having subtracted the uncertainty mass. Subsequently, the disbelief mass is computed trivially as the remaining belief mass.

$$b = (1 - u)^*(xi/ni)$$

$$d = 1 - (b + u)$$

For the finding in Hamlin et al. (2007), we can calculate that $b = (1 - 0.45) * \left(\frac{14}{16}\right) = 0.48$ and $d = 1 - (0.45 + 0.48) = 0.07$. In the Supplementary Material (see "Three worked examples"), the approach just explained is illustrated step by step on three diverse studies from Margoni and Surian (2018), for the reader to have concrete examples that may help understand the calculations.

Last, to model the cumulative nature of research, we parameterized the prior probability term a for a given study based on the xi/ni as calculated from other studies over the years preceding that study. In the Supplementary Material, we also graphically show how the subjective opinion changed over time with accumulating evidence either in favor of or counter to the presence of a preference for prosociality in infancy (to this aim, we used the year of publication, and excluded the unpublished data, as for these data date of publication reflected date of retrieval, and not when the research was actually conducted). Fig. S1 (see the Supplementary Material and OSF; OSF, 2023) displays a single barycentric triangle indicating how the subjective opinion changed over time (note that this approach is similar to a sequential Bayesian approach).

3.7. Graphically displaying subjective opinions

The resulting subjective opinions for all the findings inserted in the meta-analysis are listed in Table S1 (see the Supplementary Material). They can be visualized in two ways: as barycentric triangles (Fig. 2) and as beta distributions (Fig. 3). These are complementary ways to display the same results. Fig. 2 shows that most of the findings tend to distribute on the right part of the triangle (expressing belief) but also on the upper part, expressing a high level of uncertainty. The same conclusion can be drawn from Fig. 3 by looking at how most of the distributions are spread, reflecting again the high uncertainty of the findings.

3.8. Fusing subjective opinions

The power and the usefulness of the subjective logic approach reside not only in the fact that uncertainty can be modelled directly in the opinion regarding individual research findings. This approach can be used to aggregate or fuse together different opinions, each with its own level of uncertainty. To aggregate the subjective opinions generated from the findings meta-analyzed by Margoni and Surian (2018), we used the Weighted Belief Fusion (WBF) operator, as defined in Definition 4 (Section 2.4). In the Supplementary Materials (see "Illustration of the WBF Operator") we provide a concrete illustration of how the operator works by applying it to the three studies we also used to illustrate the process of producing subjective opinions. In short, WBF computes the fused subjective opinion $\omega_X^{\hat{\diamond}\mathbb{A}}$ (where $\mathbb{A}$ is here a set of findings or experiments on X, where each $A \in \mathbb{A}$ produces an opinion ω_X^A) by attributing more weight to opinions that carry less uncertainty. Although we use the WBF to illustrate how real findings can be aggregated, there are numerous alternative operators (e.g., the Cumulative Fusion operator does what the WBF does with the exception that every additional piece of evidence can only increase the fused belief level; for a broader discussion of all the operators, see Jøsang, 2016).

3.9. Fusing the meta-analysis studies

Using WBF operator to fuse all the findings in the meta-analysis resulted in the following fused subjective opinion ($b_X^{\hat{\diamond}\mathbb{A}} = 0.42$, $d_X^{\hat{\diamond}\mathbb{A}} = 0.20, u_X^{\hat{\diamond}\mathbb{A}} = 0.38$) where $\mathbb{A}$ is here precisely the set of findings meta-analyzed by Margoni and Surian (2018) about X, that is, the belief that infants and toddlers show a preference for prosociality over antisociality.

The result of this belief fusion operation can be visualized in Fig. 4. Panel (a) displays the fused subjective opinion as a barycentric triangle, whereas panel (b) displays the corresponding beta distribution. For this fused opinion, the projected probability (as previously defined) is calculated to be P = 0.70. It is interesting to compare this probability with the meta-analysis estimate which was reported to be P = 0.68 (or P = 0.64 when adjusted for the publication bias, see Margoni & Surian, 2018). The projected probability is slightly higher than the probability estimated by the meta-analysis. Classical approaches can provide an indication of the level of uncertainty too; one can simply look at the CI at n%. However, the subjective logic approach provides a level of uncertainty calculated by explicitly reasoning about what factors are likely to influence the estimate and in which way. In our running example, level of uncertainty (as a product of the four selected factors) can be judged as relatively high, given the value of $u_X^{\hat{\diamond}\mathbb{A}}$ (0.38). The broad arc of the beta distribution displayed in Fig. 4b indeed does not allow us to discern a particularly distinctive peak. To reiterate, the strength of this approach is that different sources of uncertainty (confidence intervals, possible moderators, etc.) are directly encoded in the uncertainty value, which is together with the beta distribution everything we need to interpret the results of a given body of research.

We can also reflect on the role of the prior probability parameter in determining the fact that the projected probability (0.70) is slightly higher than the estimate from the meta-analysis (0.68 or 0.64, if adjusted for the publication bias). This discrepancy can be partially explained by observing that a-priori probabilities were calculated from the results of previous studies reported at any given point, and fused opinions were thus swayed by initial findings which tended to be particularly favorable (e.g., Hamlin et al., 2007). An alternative approach would be to use a uniform 0.50 a-priori probability for each finding. As discussed above, the choice of approach for establishing prior probabilities is flexible and can be tailored to suit the context. Such a change would result in a projected probability of 0.61, more in line with the (publication-bias adjusted) 2018 meta-analysis (for a graphical display see Figs. 4c and 4d). However, to reiterate, the strength of subjective logic in this context is not the projected probability value, but the possibility to reason explicitly about uncertainty.

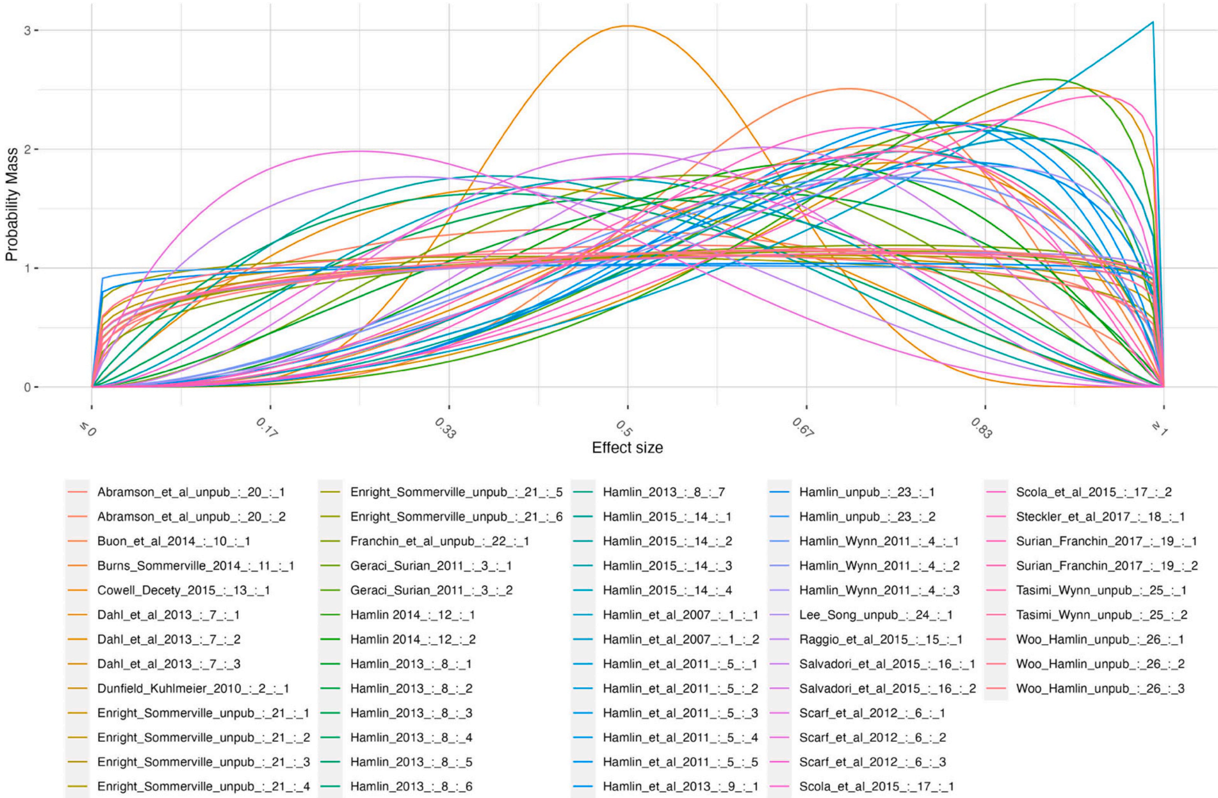

Fig. 3. *Subjective Opinions Visualized as Beta Distributions for the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis. Note*. Numbers next to the studies in the legend refer to the ID and the Study ID respectively as these can be found in the dataset (please see the dataset on OSF; OSF, 2023).

3.10. Fusing subjective opinions separated by group

A further possibility would be to fuse opinions separated by different groups of findings. The process of encoding uncertainty, modelling subjective opinions and fusing them together is the same as we already have illustrated. However, we would now be able to address questions such as 'Do older infants and toddlers as a group show a stronger preference for prosociality than younger infants?', 'Is the preference for helpers stronger or weaker than the preference for fair distributors, or than the one for agents who give as opposed to take?', or 'Does asking children to manually reach for one of the agents elicit more preferences for prosociality than asking children to selectively help one of the agents?'. For the first question, we compared groups of participants; for the second the effectiveness of different stimuli (scenarios presented to children); and for the third the effectiveness of different dependent variables (that still tap on the same phenomenon).

The visualization of the resulting fused subjective opinions is displayed in Fig. 5, Fig. 6, and Fig. 7 for the first, the second, and the third question respectively. The general message is that subjective logic allows us to aggregate different sources of uncertainty, and it does so by explicitly modelling uncertainty in the subjective opinion. Fused opinions can then be compared as long as (a) the aim of the studies being fused and compared is broadly the same (e.g., answering to the question whether infants have a moral sense), and (b) the (eventually) different effect size indexes and dependent variables are all mapped to a single generic probability interval 0–1.

To reiterate, the strength and usefulness of subjective logic is that it allows to *explicitly* reason about uncertainty. Take Fig. 5 and the comparison between younger and older children (here, to parse infants into younger and older we simply followed the criterion used in Margoni and Surian, 2018). Not only it can be observed that the two groups share a similar tendency with respect to the preference for prosociality (a point that can also be made by comparing the projected probability for the fused opinions regarding younger and older children, 0.71 and 0.69 respectively[8]), but it can also be observed that both sets of findings are characterized by high degrees of uncertainty, as evidenced by the flat and broad arcs of the corresponding beta distributions (see Fig. 5b). Similar observations can be made for Figs. 6 and 7, which show similar patterns of results across scenarios (Fig. 6) or dependent variables (Fig. 7), and show that,

[8] Another possibility would of course be to compare belief, disbelief, and uncertainty values of the two fused subjective opinions, which in the specific case are $(b = 0.41, d = 0.19, u = 0.40)$ for the opinion about the younger children, and $(b = 0.43, d = 0.22, u = 0.35)$ for the opinion about the older children. An even more nuanced approach would be to make use of operators such as the Unfusion or Fission operator, specifically designed to compare different fused subjective opinions.

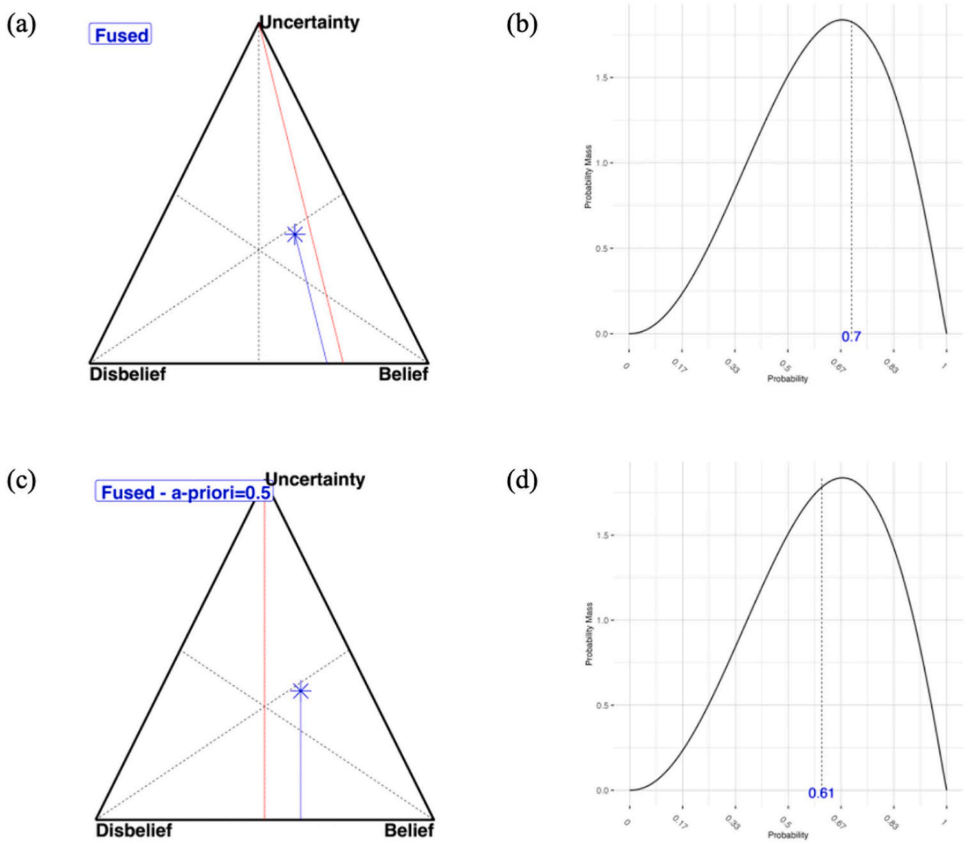

Fig. 4. *Barycentric Triangle and Beta Distribution Visualization of the Fused Subjective Opinion from the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis*. *Note*. Panel (a) displays the barycentric triangle visualization of the aggregation of all the subjective opinions modelled after the findings inserted in the meta-analysis by Margoni and Surian (2018) into a single fused subjective opinion (the blue line indicates the projected probability – see Definition 2, Section 2.2 – and the red one indicates the prior probability *a*). Panel (b) displays instead its corresponding beta distribution (the vertical, dashed line indicates where the projected probability for the fused subjective opinion lies). Panels (c) and (d) display the same results with the exception that the calculation was made by using a uniform 0.50 a-priori probability for each finding (rather than using a cumulative approach).

independently of the set of findings, opinions are fraught with uncertainty.

Last, not only it is possible to fuse opinions together to compare different groups of findings from experiments using the same measure (e.g., forced-choice preferential measure), but it is also possible both to compare fused opinions from groups of experiments employing different methodological paradigms and to fuse together all these opinions in a single fused opinion. In the Supplementary Material, we provide an example of how this can be done by fusing together the findings from the meta-analyzed experiments which employed the preferential reaching paradigm and a set of findings from experiments using the Violation-of-Expectation (VOE) paradigm (for an overview, see Margoni, Surian, & Baillargeon, 2024) to test for the presence in infants of the expectation that individuals behave in accordance with the moral principle of fairness (Sloane, Baillargeon, & Premack, 2012). Both methods and bodies of research (preferential reaching tasks and VOE looking time tasks) are used to argue that humans possess an early-emerging moral sense, and subjective logic would allow researchers to encode both of them in a single fused subjective opinion expressing the degree to which we believe in such a conclusion about human nature, given the uncertainty surrounding the evidence at our disposal.

3.11. Summary of the findings

By taking advantage of a set of findings already meta-analyzed, we have illustrated the potential of subjective logic to allow researchers to reason explicitly about uncertainty. We have showcased how uncertainty can be encoded and aggregated in a precise manner in a high-level subjective opinion that ultimately addresses the question about whether infants display a preference for prosociality over antisociality. We have shown how findings can be expressed as low-level subjective opinions, following a two-phase process that first populates the opinion with an uncertainty value, which can be derived from various sources of uncertainty and error, and then provides belief and disbelief values. Next, low-level opinions were fused by using the WBF operator, and the resulting fused subjective opinion was displayed as a barycentric triangle and a beta distribution (see Fig. 4), revealing high level of uncertainty. Last, three purely illustrative comparisons between fused opinions based on separated subsets of the meta-analyzed dataset revealed that a

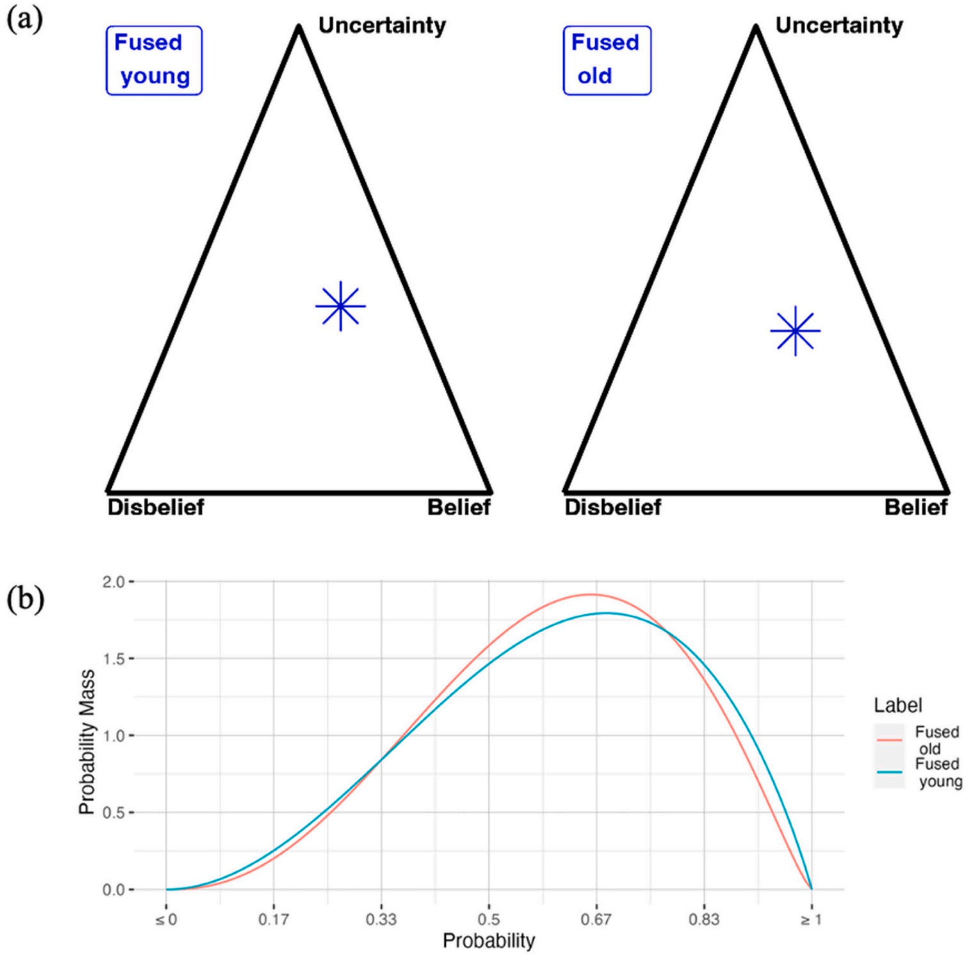

Fig. 5. *Fusion of the Subjective Opinions from the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis into Young and Old Age Groups. Note.* Panel (a) displays the barycentric triangle visualization of the aggregated subjective opinions of experiments conducted with younger infants (mean age ≤ 12 months, 29 days; following Margoni and Surian, 2018 criterion) and older infants, respectively. Panel (b) displays the corresponding beta distributions for these two fused subjective opinions (younger infants in green, older infants in red).

high degree of uncertainty continues to characterize the findings, far beyond possible differences in belief values and projected probabilities between the subsets. Applied to our example, the subjective logic approach revealed that the meta-analytic estimate that about two out of three infants prefer prosociality over antisociality and the research area in which these findings had been appearing are surrounded by extremely high levels of uncertainty. Indeed, recall that the level of uncertainty (0.38) was approximately equal to the level of belief (0.42) in the statement that infants prefer prosocial agents, with the remaining 0.20 expressing disbelief. And it is exactly because of this observation that we are left wondering about the reality of the effect originally reported by Hamlin et al. (2007). In sum, the subjective logic approach helped us expose a high degree of uncertainty, reason about it in an explicit way (all sources of uncertainty can be easily and clearly identified), and incorporate uncertainty in a fused opinion expressing levels of belief, disbelief and uncertainty about the statement that infants prefer prosociality over antisociality.

4. Discussion

Findings (and conclusions based on them) are always to some extent uncertain. Subjective logic has been applied to decision making under uncertainty. In this paper, we argue that uncertainty can and should be explicitly considered as a primary aspect of any empirical enterprise, and its assessment should inform the interpretation of the data. This process of interpretation of the findings can ultimately be framed as a process of decision making under (sometimes extreme) uncertainty. Indeed, addressing any research question based on actual data requires a degree of inference which is inherently affected by uncertainty. To this end, subjective logic promises to be a powerful tool complementing meta-analysis and classical approaches to statistics (not substituting them) to explicitly compute and encode various sources of uncertainty into the reasoning about the findings.

There is of course an undisputed consensus on the importance of uncertainty for the interpretation of research data, and classical approaches to statistics provide numerous ways of reasoning about uncertainty and heterogeneity among findings. However, what

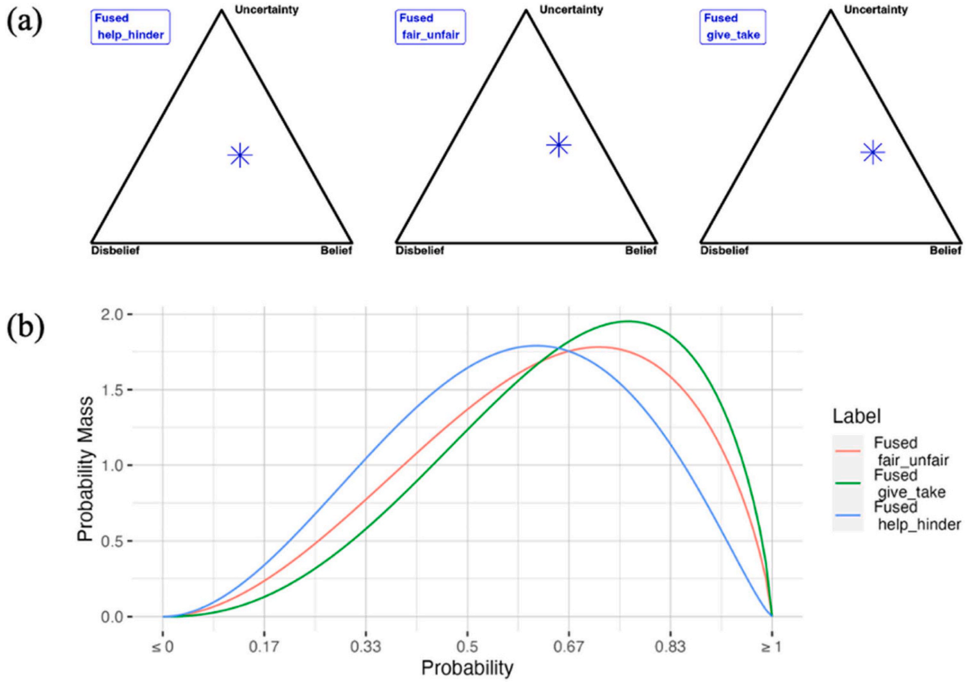

Fig. 6. *Fusion of the Subjective Opinions from the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis into Scenario Groups***.** *Note*. Panel (a) displays the barycentric triangle visualization of the aggregated subjective opinions of experiments employing help/hinder scenarios, fair/unfair scenarios, and give/take scenarios, respectively. Panel (b) displays the corresponding beta distributions for these three fused subjective opinions (a different color is used for each fused opinion).

subjective logic can uniquely offer is a way to directly encode uncertainty in the answer provided to the given research question, where the uncertainty estimate is the result of a transparent and explicit operation, that is, the result of a calculation which should be openly discussed and eventually revised according to what the research community thinks. After a meta-analysis is conducted, subjective logic can help initiating a careful but explicit reasoning about what is likely to create uncertainty in the research findings, and how – which is the beauty, the complexity, and the usefulness of the approach we are putting forward.

We have showcased how this approach can be readily used to re-analyze findings relevant for psychologists to assess the high-level question of whether children show early in life an initial sensitivity to moral scenarios. The novelty with respect to a meta-analytic approach is that heterogeneity among studies or findings is not addressed with a set of moderation analyses but can be modelled and visualized directly and explicitly in the main analysis output. Researchers are thus provided with a powerful tool for the often-complex task of interpreting cases of discrepant findings (such as failed attempts to replicate) or inconclusive bodies of research. In general, this approach allows researchers to reason in a precise manner about the uncertainty pertaining to findings, and how various sources of uncertainty affect the conclusions that can be drawn from a given set of findings. Aggregated subjective opinions are answers to research questions that explicitly incorporate and express (un)certainty. For instance, once analyzed through the lens of subjective logic, it becomes apparent that the body of research on infants' preference for prosociality carries a high degree of uncertainty, which influences the confidence in the conclusion that infants, indeed, show a proclivity to prefer prosocial over antisocial agents. Moreover, all the sources of uncertainty are clearly visible in the model, and the decisions about which sources (of both first- and second-order uncertainty) to include and how to formalize penalty scores are available to the research community for further scrutiny and assessment. In principle, anything that researchers may believe would reduce our confidence in the results can be modelled as a source of uncertainty.

4.1. Limitations: the role of flexibility

Uncertainty is a primary consideration in research, and researchers need tools favouring a *precise, explicit,* and *transparent* process of reasoning about uncertainty and error in the context of data analysis and interpretation. Subjective logic allows to explicitly factor in various sources of uncertainty and aggregate data irrespective of the sources of uncertainty affecting the results. These possibilities come however with an important caveat. The very core process of defining belief, disbelief, and – above all – uncertainty is necessarily *subjective*. The process reflects the entire set of decisions made by who analyzes the data (as we illustrated in our running example). Thus, the limitation of this approach is that different readers and analysts may interpret the experimental setting differently. They may choose different ways to encode the findings, which may disagree. The same experimental setting and data could yield different interpretations. To an extent this situation simply reflects the nature of scientific discourse, and it is possible that the formulation of

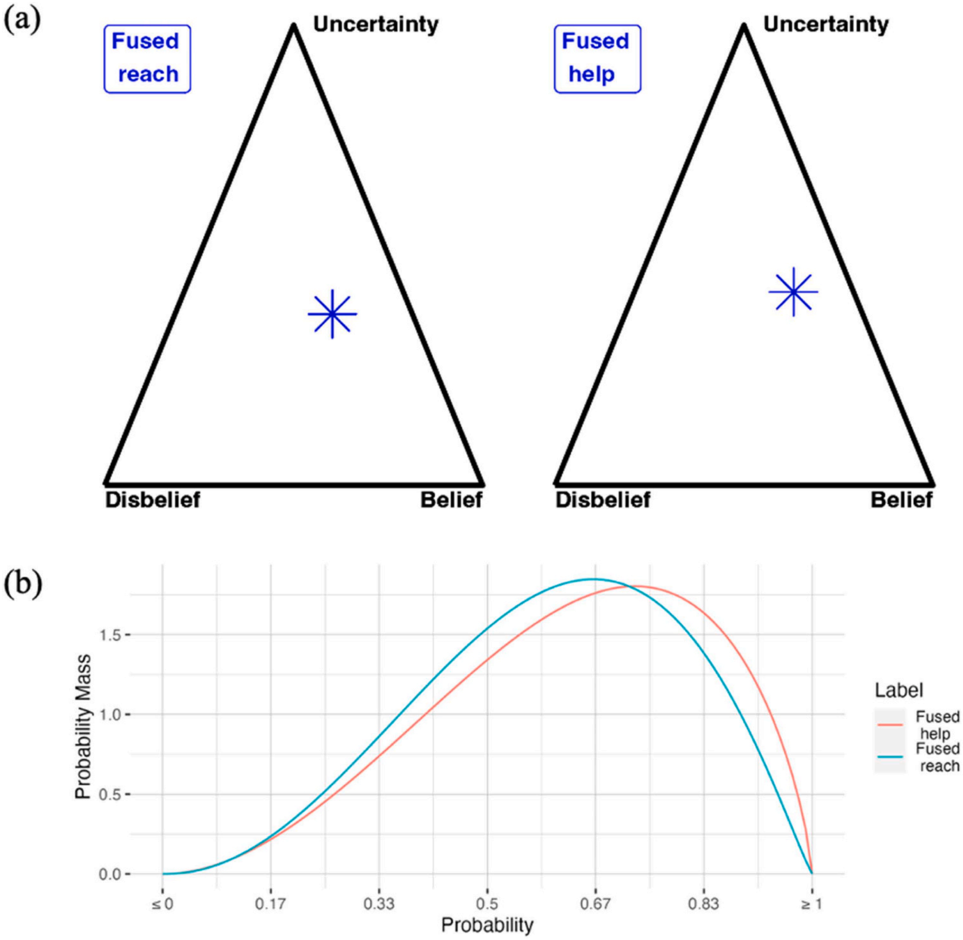

Fig. 7. *Fusion of the Subjective Opinions from the Findings Inserted in* Margoni and Surian (2018) *Meta-analysis into Reach and Help Measure Groups*. *Note*. Panel (a) displays the barycentric triangle visualization of the aggregated subjective opinions of experiments employing manual reaching and selective helping paradigms, respectively. Panel (b) displays the corresponding beta distributions for these two fused subjective opinions (manual reaching in green, selective helping in red).

subjective opinions can be standardized to some extent. However, the process of capturing reasoning about experiments and data we have here introduced draws an explicit line between method, data, and conclusions. It can serve to explain an interpretation of a group of experiments that captures the nuances of how results are affected by different sources of uncertainty. Whereas uncertainty is unavoidable and second-order uncertainty can be inherently subjective, we are left with two options. We could leave the various sources of uncertainty that influence results implicit and unacknowledged, or we could enable researchers to formalize their interpretation of the underlying first-order and second-order uncertainties so that results can be interpreted through this lens. Following this second path certainly would not offer unquestionable interpretations. Quite the opposite. It would produce subjective opinions and disagreement between researchers. However, subjective logic offers a formal basis by which to present and aggregate results in a way that makes these interpretations *explicit*. Thus, the strength of this approach is that it allows to explicitly reason about first-order and second-order uncertainty, unavoidable aspects of research. And the higher the levels of uncertainty, the more useful and important it becomes to complement meta-analysis with a subjective logic approach.

Our work was purely illustrative and introductory, as we merely aimed to provide the reader with the basics to understand and value the process of encoding empirical findings as subjective opinions and fusing them together. However, when applying this approach to given data, researchers might want to consider preregistering all the decisions implied in the process of modelling subjective opinions. It is somehow in the nature of a subjective process of decision making that in order to be shared and discussed with others it should be communicated clearly and *transparently*. Transparency is of paramount importance and can be achieved by providing a clear description of how subjective opinions were modelled and subsequently fused (the source code can be made fully available and easily interpretable). Moreover, being subjective must not be confused with being decided by a single person or small group, rather such a subjective process could be grounded in a collective agreement which could be established through the use of surveys to be circulated among researchers in a given field. Ultimately, what began as subjective can later become objective in the sense of commonly accepted within a given community. Subjective logic is thus flexible as it is not tied to specific a priori criteria and parameters, but once the parameters are discussed and agreed upon within the community, they can be applied in a fixed and consistent manner. Examples of such practices within research communities are attempts to reach consensus over the definition of key concepts (for a relevant and recent example within psychology research, see Quesque et al., 2024) or consensus papers seeking to identify best methods and results.

The question of whether to accept or reject a hypothesis or a theoretical claim can be further informed by an open discussion about (a) how uncertainty likely biases the conclusion to be drawn, and (b) how uncertainty has to be understood in the given context, that is, encoded and modelled into the answer the community eventually agrees to give to a research question. Issues regarding the definition and parametrization of uncertainty values are not impossible to be addressed by a community as a whole. We may recall the lesson about the 'wisdom of the crowds' given by Francis Galton at the beginning of the last century, who showed how about 800 villagers interviewed about the weight of an ox reached *on average* a nearly perfect estimate (1197 approximating the real weight of 1198 pounds) while at the same time disagreeing among themselves to a certain extent (no one got the exact weight; Galton, 1907). Favoured by transparency, open discussion about the best criteria used and decision made to analyze data and, here, to model subjective opinions is an entirely viable opportunity, and one that can contribute to research progress. The subjective logic approach gives researchers an opportunity to reason explicitly about factors affecting the interpretation of a given body of research that are not implicit and buried in the data somehow, but rather connected with our opinions about *how* the data were collected. Making this information explicit is by itself a valid and substantial contribution to the interpretation of a body of research.

5. Conclusion

In conclusion, we have illustrated the potential and feasibility of subjective logic to complement meta-analysis in attaining the goal of capturing in a precise manner uncertainty affecting the findings. By treating research findings as subjective opinions, it becomes possible to incorporate explicitly, flexibly, and transparently uncertainty in the process of providing an answer to a research question. Given the ubiquitousness and centrality of error and uncertainty in research, subjective logic promises to be an interesting and particularly powerful tool.

Financial support

This work was supported by the EPSRC CITCOM grant (EP/T030526/1) to N. W.

CRediT authorship contribution statement

Francesco Margoni: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Neil Walkinshaw:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation.

Data Availability

All data, analysis code, and supplementary materials are available on the Open Science Framework: https://osf.io/w9f8a/.

Acknowledgments

We would like to thank Martin Shepperd and Alvin Wei Ming Tan for their helpful comments on a previous version of this manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.infbeh.2024.101978.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108–119. https://doi.org/10.1002/per.1919

Blake, K. R., & Gangestad, S. (2020). On attenuated interactions, measurement error, and statistical power: Guidelines for social and personality psychologists. *Personality and Social Psychology Bulletin, 46*, 1702–1711. https://doi.org/10.1177/0146167220913363

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H., Adem, M., Adriaans, J., & Van Assche, J. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences, 119*, Article e2203150119. https://doi.org/10.1073/pnas.2203150119

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*, 637–644. https://doi.org/10.1038/s41562-018-0399-z

Cerutti, F., Kaplan, L. M., Norman, T. J., Oren, N., & Toniolo, A. (2015). Subjective logic operators in trust assessment: an empirical study. *Information Systems Frontiers, 17*, 743–762. https://doi.org/10.1007/s10796-014-9522-5

Chambers, C. D. (2013). Registered reports: A new publishing initiative at cortex. *Cortex, 49*, 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*, 7–29. https://doi.org/10.1177/0956797613504966
Davis, J., Redshaw, J., Suddendorf, T., Nielsen, M., Kennedy-Constantini, J., Oostenbroek, & Slaughter, V. (2021). Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Psychological Science, 16*, 1373–1397. https://doi.org/10.1177/1745691620959834
Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety, 31*, 105–112. https://doi.org/10.1016/j.strusafe.2008.06.020
Dempster, A. P. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological, 30*, 205–232. https://doi.org/10.1111/j.2517-6161.1968.tb00722.x
Dietzel, S., van der Heijden, R., Decke, H., & Kargl, F. (2014). A flexible, subjective logic-based framework for misbehavior detection in V2V networks. *In Proceeding of the International Conference on Information Fusion*, 1–6. https://doi.org/10.1109/wowmom.2014.6918989
Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy, 22*, 421–435. https://doi.org/10.1111/infa.12182
Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature, 75*, 450–451. https://doi.org/10.1038/075450a0
Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine, 8*. https://doi.org/10.1126/scitranslmed.aaf5027
Grand, J. A., Rogelberg, S. G., Banks, G. C., Landis, R. S., & Tonidandel, S. (2018). From outcome to process focus: Fostering a more robust psychological science through registered reports and results-blind reviewing. *Perspectives on Psychological Science, 13*, 448–456. https://doi.org/10.1177/1745691618767883
Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*, 557–559. https://doi.org/10.1038/nature06288
Ioannidis, J. P. (2012). Scientific inbreeding and same-team replication: Type D personality as an example. *Journal of Psychosomatic Research, 73*, 408–410. https://doi.org/10.1016/j.jpsychores.2012.09.014
Jøsang, A. (2001). A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 9*, 279–311. https://doi.org/10.1016/s0218-4885(01)00083-1
Jøsang, A. (2016). Subjective logic: A Formalism for Reasoning Under Uncertainty. Springer.
Jøsang, A., & Bondi, V. A. (2000). Legal reasoning with subjective logic. *Artificial Intelligence and Law, 8*, 289–315. https://doi.org/10.1023/a:1011219731903
Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*, 355–362. https://doi.org/10.31234/osf.io/97gpc
Lucca, K., Capelier-Mourguy, A., Cirelli, L., Byers-Heinlein, K., Dal Ben, R., Frank, M.C., … & Hamlin, K. (2021). Infants' social evaluation of helpers and hinderers: A large-scale, multi-lab, coordinated replication study. Pre-print retrieved at https://doi.org/10.31234/osf.io/qhxkm.
Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537–542. https://doi.org/10.1177/1745691612460688
Margoni, F., Block, K., Hamlin, K., Zmyj, N., & Schmader, T. (2023). Meta-analytic evidence against sex differences in infants' and toddlers' preference for prosocial agents. *Developmental Psychology, 59*, 229–235. https://doi.org/10.1037/dev0001421
Margoni, F., & Shepperd, M. (2020). Changing the logic of replication: A case from infant studies. *Infant Behavior and Development, 61*, Article 101483. https://doi.org/10.1016/j.infbeh.2020.101483
Margoni, F., Surian, L., & Baillargeon, R. (2024). The Violation-of-Expectation paradigm: A conceptual overview. *Psychological Review, 131*, 716–748. https://doi.org/10.1037/rev0000450
Margoni, F., & Surian, L. (2018). Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Developmental Psychology, 54*, 1445–1455. https://doi.org/10.1037/dev0000538
Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34*, 103–115. https://doi.org/10.1086/288135
Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16*, 406–419. https://doi.org/10.1037/a0024377
Munder, T., Brutsch, O., Leonhart, R., Gerger, H., & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review, 33*, 501–511. https://doi.org/10.1016/j.cpr.2013.02.002
Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*, 1422–1425. https://doi.org/10.1126/science.aab2374
Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published reports. *Social Psychology, 45*, 137–141. https://doi.org/10.1027/1864-9335/a000192
Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science, 349, aac4716. https://doi.org/10.1126/science.aac4716.
Oren, N., Norman, T. J., & Preece, A. (2007). Subjective logic and arguing with evidence. *Artificial Intelligence, 171*, 838–854. https://doi.org/10.1016/j.artint.2007.04.006
OSF (2023). *Subjective logic to transparently address uncertainty in research findings*. https://osf.io/w9f8a/.
Quesque, F., Apperly, I., Baillargeon, R., Baron-Cohen, S., Becchio, C., Bekkering, H., & Brass, M. (2024). Defining key concepts for mental state attribution. *Communications Psychology, 2*, 29. https://doi.org/10.1038/s44271-024-00077-6
Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PloS One, 10*, Article e0140570. https://doi.org/10.1371/journal.pone.0140570
Schlingloff, L., Csibra, G., & Tatone, D. (2020). Do 15-month-old infants prefer helpers? A replication of Hamlin et al. (2007). *Royal Society Open Science, 7*, Article 191795. https://doi.org/10.1098/rsos.191795
Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods, 22*, 322–339. https://doi.org/10.1037/met0000061
Shafer, G. (1976). A mathematical theory of evidence. Princeton University Press.
Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science, 23*, 196–204. https://doi.org/10.1177/0956797611422072
Stanley, D., Jarrell, S. B., & Doucouliagos, H. (2010). Could it be better to discard 90% of the data? A statistical paradox. *The American Statistician, 64*, 70–77. https://doi.org/10.1198/tast.2009.08205
Szucs, D., & Ioannidis, J. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature (Article) *PLoS Biology, 15*, Article e2000797. https://doi.org/10.1371/journal.pbio.2000797.
Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science, 9*, 661–665. https://doi.org/10.1177/1745691614552498
Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., … Nosek, B. A. (2019). Scientific utopia III: Crowdsourcing science. *Perspectives on Psychological Science, 14*, 711–733. https://doi.org/10.1177/1745691619850561
Van Der Heijden, R. W., Kopp, H., & Kargl, F. (2018). *Multi-source fusion operations in subjective logic*. In Proceedings of. *the International Conference on Information Fusion*, 1990–1997. https://doi.org/10.23919/icif.2018.8455615
van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004
Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 632–638. https://doi.org/10.1177/1745691612463078

Walkinshaw, N., & Hierons, R. (2023). Modelling second-order uncertainty in state machines. *IEEE Transactions on Software Engineering*. https://doi.org/10.1109/tse.2023.3250835

Walkinshaw, N., & Shepperd, M. (2020). Reasoning about uncertainty in empirical results. *In Proceedings of the Evaluation and Assessment in Software Engineering*, 140–149. https://doi.org/10.1145/3383219.3383234

Wilson, B. M., Harris, C. R., & Wixted, J. T. (2020). Science is not a signal detection problem. *Proceedings of the National Academy of Sciences, 117*, 5559–5567. https://doi.org/10.1073/pnas.1914237117