# Introducing the Choice-Confidence (CHOCO) Model for Bimodal Scales Data: Application to the Relationship between Reality Beliefs about AI-Generated Faces and Beauty

Dominique Makowski[1,2], Ana Neves[1], and Andy Field[1]
[1]School of Psychology, University of Sussex
[2]Sussex Centre for Consciousness Science, University of Sussex

TO DO.

*Keywords:* subjective scales, choice confidence model, reality beliefs, AI-generated faces, attractiveness

## Introduction

Despite significant advancements in psychological science following the replication crisis (Collaboration, 2015), its progress is still hindered by its sub-optimal (or inappropriate) usage of statistical tools (Makowski & Waggoner, 2023). A prevalent issue is the continued reliance on linear models that assume normally distributed (Gaussian) data - as this assumption often does not hold true for many types of psychological outcomes. For instance, reaction times typically exhibit skewed distributions, choices can be represented as binary variables, and count data consists of strictly positive integers. Applying models that presume normality and model the "mean" of the outcome variable can lead to misinterpretation and potentially misleading conclusions when applied indiscriminately. It is thus important that psychologists use models that can best describe (or generate) the data they collect, to fully exploit them and bring more nuance and accuracy to their conclusions.

Among the most commonly collected data in psychology are responses on subjective scales, such as Likert-type items or visual analog scales, which exhibit some fundamental properties: these responses are bounded (and can be rescaled to a 0-1 range) and frequently display clustering at the extremes. Traditional linear models being ill-suited for such data, researchers have turned to using Beta distributions to model this data (instead of Gaussian), suited for continuous data within the (0,1) interval (i.e., excluding extreme responses). To address the frequent occurrence of exact zeros and ones (i.e., extreme values), zero-one inflated beta (ZOIB) models have been developed (Ospina & Ferrari, 2012) to accommodate the excess of boundary values by incorporating additional components that model the probabilities of responses at 0 and 1 as a separate, independent process.

### The Beta-Gate Model

The Beta-Gate model is a reparametrized Ordered Beta model (Kubinec, 2023)[1] available in the *cogmod* package in R (https://github.com/DominiqueMakowski/cogmod), in which participants' answers on bounded scales are conceptualized as latent responses that can fall past a pair of probabilistic "gates" (or cutpoints) that control whether the response is recorded as an extreme (0 or 1) or as a nuanced, continuous value in between (Figure 1). These distance of these gates from the edges of the scale varies based on two interpretable parameters: *pex* (the propensity by which people are likely to

Dominique Makowski
Ana Neves
Andy Field

This preprint is a non-peer-reviewed work from the **Reality Bending Lab**.

Author roles were classified using the Contributor Role Taxonomy (CRediT; https://credit.niso.org/) as follows: Dominique Makowski: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft; Ana Neves: Data curation, Writing – original draft, Writing – review & editing; Andy Field: Writing – original draft, Writing – review & editing

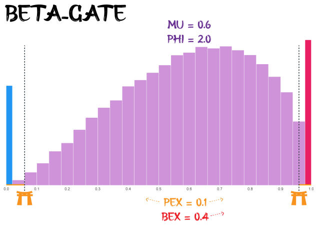
Correspondence concerning this article should be addressed to Dominique Makowski, Email: D.Makowski@sussex.ac.uk

[1] In the Ordered Beta model, the cutpoints on the log-scale are directly used as parameters, instead of being derived from *pex* and *bex*.

answer extreme values), and *bex*, a bias toward the upper extreme (1) versus the lower (0). A person's internal response that lies close to the edge might be "caught" by a gate and recorded as an extreme, while others pass through to express a continuous response (Beta-distributed with $\mu$ (*mu*) and $\phi$ (*phi*) as its mean and precision parameters). The Beta-Gate model is based on the idea that extreme values can emerge not just from a fundamentally different underlying processes - as assumed in ZOIB models - but from a common process governed by thresholds of decisiveness and confidence.

**Figure 1**

*The Beta-Gate Distribution is a reparametrized ordered Beta model (Kubinec, 2023) that is governed by 4 parameters. 'Mu' and 'phi' correspond to the mean and precision of the continuous part of the distribution (between 0 and 1), and 'pex' (propensity of extremes) and 'bex' (balance of extremes) indirectly control the proportion of zeros and ones by specifying the location of the "gates", past which the latent response process is likely to generate extreme values. Specifically, 'pex' defines the total distance of both gates from the extremes (in yellow), and 'bex' determines the proportion of the right gate distance relative to the left. In this example, the total distance from the extremes is 'pex' = 0.1, with 40% ('bex' = 0.4) of that distance being on the right (and 60% on the left). The left gate is thus located at 0.6, and the right at 1-0.04 = 0.96.*



Mathematically, the Beta-Gate distribution defines the observed outcome $x \in [0, 1]$ as a mixture of three components; a point mass at 0, a point mass at 1, and a continuous Beta density over $x \in (0, 1)$, scaled by the remaining probability mass. The probability of these components are:

- $P(x = 0) = \text{logistic}\left(\text{logit}(pex \cdot (1 - bex)) - \text{logit}(\mu)\right)$
- $P(x = 1) = 1 - \text{logistic}\left(\text{logit}(1 - pex \cdot bex) - \text{logit}(\mu)\right)$
- $P(x \in (0, 1)) = 1 - P(x = 0) - P(x = 1)$

The continuous part follows a Beta distribution with parameters[2]:

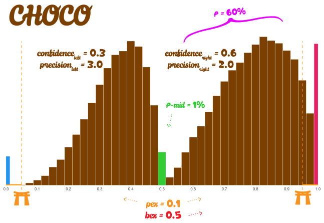
$$Beta(\alpha = \mu \cdot 2\phi, \quad \beta = (1 - \mu) \cdot 2\phi)$$

## The Choice-Confidence (CHOCO) Model

Decision-making is often conceptualized as involving distinct processes: the choice itself, and the confidence associated with that choice. In experimental paradigms, these can be somewhat disentangled by prompting participants to make a discrete choice selection (e.g., "True" vs. "False"), followed by a separate confidence rating. However, this artificial separation makes its joint analysis difficulty, and may not reflect real-world decision-making, where individuals often express both choice and confidence simultaneously using a single, continuous scale. In such scales, each side can represent a distinct latent category, and the distance from the midpoint can indicates the level of confidence or certainty. This integrated response format typically results in bimodal distributions, with peaks corresponding to the mean confidence on either side. Traditional beta regression models, which assume unimodal distributions within the (0,1) interval, are ill-suited for such data. One alternative is to transform the data into two variables a posteriori: binarizing the side to represent choice and calculating the absolute distance from the midpoint to represent confidence. These can then be modeled separately, for instance, using logistic regression for choice and beta regression for confidence (see Makowski et al., 2025 for an example). While this approach can provide additional insights into underlying mechanisms compared to a unique model, it assumes psychological and statistical independence between choice and confidence, which may not hold true in practice.

**Figure 2**

*The CHOCO Model uses a mixture of Beta-Gate distributions to model separately the right and left sides of the scale (e.g., a rating of whether a statement was 'Truth' vs. 'Lie'), as well as their relative proportion. In this example, the participants are more likely overall (p = 60%) to select the right side of the scale ('Lie') than the left ('Truth'). They are also more confident in their choice (confright = 0.6 vs. confleft = 0.3). Extreme values (zeros and ones) are governed by the same mechanism as for Beta-Gate models.*



To model data of subjective scales in which the left and right sides can be conceptualized as two different choices (e.g., True/False, Agree/Disagree, etc.) and the magnitude of the response (how much the cursor is set away from the midpoint) as the confidence, we introduce the Choice-Confidence (CHOCO) model (Figure 2). It consists of a three-part mix-

[2] Note that *phi* is scaled in Beta-Gate models relative to the traditional mu/phi Beta particularization so that a phi of 1 corresponds to a uniform distribution - to facilitate setting priors on this parameter

ture on $x \in [0, 1]$:

- An (optional) point-mass at the midpoint *mid* (typically 0.5) of weight $p_{\text{mid}}$ for undecided or neutral responses.

- A left-choice component governed by a Beta-Gate density on the rescaled variable $x/mid$ with mean *1 - confleft*, precision *precleft*, and boundary-excess parameter *pex(1 - bex).*

- A right-choice component governed by a Beta-Gate density on the rescaled variable $(x - mid)/(1 - mid)$ with mean *confright*, precision *precright*, and boundary-excess parameter *pex x bex*.

The overall probability of the right choice (relative to the left choice) is controlled by a main parameter $p$. The full CHOCO density is:

$$\begin{cases} pmid, & x = mid, \\ (1 - pmid)(1 - p)\,\dfrac{1}{mid} \cdot \text{BetaGate}\left(\dfrac{x}{mid}\right), & 0 < x < mid, \\ (1 - pmid)\,p\,\dfrac{1}{1 - mid} \cdot \text{BetaGate}\left(\dfrac{x - mid}{1 - mid}\right), & mid < x < 1. \end{cases}$$

By coupling choice probability $p$, midpoint mass $p_{\text{mid}}$, and side-specific Beta-Gate parameters ($conf$, $prec$, $pex$, $bex$), CHOCO flexibly captures both bimodality and confidence intensity in a single unified model. Despite this theoretical appeal, it is unclear whether this heavily parametrized model can be estimated reliably from data, and whether it can provide more useful insights than simpler alternatives.

## Aim of the Present Study

**Study 1** aims to evaluate the CHOCO model's ability to better capture subjective scale responses that (potentially) reflect an underlying discrete choice, in comparison to existing models such as the ZOIB and Beta-Gate. Specifically, we will assess whether 1) CHOCO provides improved model fit, 2) yields deeper insights into population-level effects than traditional approaches (gender differences in reality beliefs), and 3) allows for the reliable estimation of interpretable individual-level parameters through random effects.

To this end, we analyze data from two separate studies in which participants judged whether a face image was AI-generated ("fake") or a real photograph. **Study 2** will apply this model to more subtle effects, such as the effect perceived facial attractiveness and beauty on reality judgments.

# Study 1

**TODO**: Write some introduction about reality beliefs in the context of ambiguity. Underline that what drives judgments and beliefs of reality remains unclear. And this is important in today's post-truth era.

In a post-truth era (Lewandowsky et al., 2017), where near-perfect simulations are increasingly feasible due to rapid technological advancements, distinguishing between what is real and what is fake poses a growing challenge [**REF**]. As these technologies advance and physical cues become indistinguishable, individuals are likely to rely more heavily on alternative sources of information -such as contextual cues and cognitive heuristic -when making reality judgments [i.e., simulation monitoring; Makowski (2018), Makowski, Sperduti, et al. (2019)]. This reliance is particularly pronounced under conditions of high ambiguity, where processing demands are higher and the atomization and decontextualization of information (especially prevalent in online environments) further complicate the discernment of authenticity.

Another unknown is whether these beliefs are fully stimuli-driven or whether there are some stable individual tendencies that are at stake. For instance, in the absence of clear contextual information, individual-level characteristics may increasingly guide beliefs about reality. Traits such as narcissism have been shown to influence such judgments (Makowski et al., 2025). Moreover, it has been proposed that the interaction between stimulus characteritics and individual differences plays a crucial role in shaping reality beliefs. For example, self-reported arousal has been found to predict the likelihood of perceiving an image as real (Azevedo et al., 2020), while judging a stimulus as fake has been associated with emotional downregulation (Makowski, Sperduti, et al., 2019). [**should probably add other examples**]

These findings underscore that what drives beliefs and judgments of reality remains unclear. Yet, understanding these mechanisms is increasingly vital given the proliferation of AI-generated content and the growing risk of misjudging the authenticity of stimuli, particularly in contexts with significant societal consequences (e.g., the influence of misinformation on democratic processes).

## Methods

### *Participants*

The sample is the same as in (Makowski et al., 2025). We included all heterosexual and bisexual (these two groups did not seem to differ based on preliminary analyses and were thus grouped to maximize power) male and female participants, for a final sample of 141 participants (Mean age = 28.4, SD = 9.0, range: [19, 66]; Sex: 47.5% females). For each participant, we included only stimuli of the opposite gender (i.e., all 89 female faces for men and 20 male faces for women).

### *Procedure*

In the first phase, participants viewed 109 neutral-expression photographs of faces (random order, display time of 500 ms) from the American Multiracial Face Database (Chen et al., 2021). After each image, participants rated the face on trustworthiness, familiarity, attractiveness, and beauty using visual analog scales. In the second phase, participants were informed that "about half of the previously seen images were AI-generated". The same faces were presented again in a new random order (same display time), followed by ratings of "reality" (whether they believed the image was fake - left anchor - or a real - right anchor). Makowski et al. (2025) showed that most stimuli yielded highly variable reality ratings, suggesting that idiosyncratic cognitive processes were at play.

### *Data Analysis*

We fitted 3 models to predict the reality ratings: a ZOIB model, a Beta-Gate model, and the CHOCO model. For all models and each parameter, the full formula was entered: $Real \sim Sex + (1|Participant) + (1|Item)$ (with Sex as the main predictor and participants and items entered as random intercepts). The models were run using *brms* (Bürkner, 2017) R package, and analyzed using the *easystats* collection of packages (Lüdecke et al., 2022). To maximize the comparability across models We used the default priors (uniform) for all models, and we ran 16 chains of 1400 iterations each on the University of Sussex High-Performance Computing (HPC) cluster.

Model comparisons were performed using the *loo* R package (Vehtari et al., 2017), which computes the Widely Applicable Information Criterion (WAIC) and estimates the Expected Log Predictive Density (ELPD) and penalizes the number of parameters. We assessed model performance by examining ELPD differences and their standard errors (SE), reporting corresponding *p*-values to determine significant differences in predictive accuracy.

For the population-level effects, we will consider significant and report (using the median of the posterior distribution) effects for which the 95% Credible Interval (CI) does not include zero (and when the probability of direction *pd* is > ~97%, Makowski, Ben-Shachar, et al., 2019). For the individual-level parameters (i.e., the random intercepts of each parameter for each participant and each item), we will first analyze their reliability using the Variance-Over-Uncertainty Ratio index (*D-vour*). This index, implemented in the *performance* package (Lüdecke et al., 2021), is inspired by recent work on mixed models reliability (Rouder & Mehrvarz, 2024; Williams et al., 2021), and corresponds to the normalized ratio of observed variability to uncertainty in random effect estimates, defined as:

$$D_{\text{vour}} = \frac{\sigma_B^2}{\sigma_B^2 + \mu_{\text{SE}}^2}$$

Where $\sigma_B^2$ is the between-group variability (computed as the SD of the random effect point-estimates) and $\mu_{\text{SE}}^2$ is the mean squared uncertainty in random effect estimates (i.e., the average uncertainty). We use as *D-vour* = 0.666 as the threshold for moderately reliable random effect estimates, which corresponds to a 2:1 ratio of between-group variance to uncertainty.

Finally, we will run a correlation analysis of the models' individual-level estimates against "empirical" (indices computed directly on the observed data), including the empirical *p* ($P(y > 0.5)$), the overall *conf* ($mean(|y - 0.5|)$), *pex* ($P(y \in [0, 1])$) and *bex* ($P(y == 1)/P(y \in [0, 1])$), assessing whether the model's estimate are in-line with easily interpretable indices.

## Results

The reproducible code and full result report are available at https://github.com/RealityBending/FictionChoco.

### *Model Comparison*

The models did converge without divergent transitions, and the effective sample size was sufficient for all parameters (all $n_{\text{eff}} > 1000$). The difference in predictive accuracy, as indexed by Expected Log Predictive Density (ELPD-WAIC), suggests that the *CHOCO* is the best model ($ELPD = -203.54$), followed by *Beta-Gate* ($\Delta_{ELPD} = -1794.57 \pm 63.12, p < .001$) and *ZOIB* ($\Delta_{ELPD} = -1833.59 \pm 63.52, p < .001$). See Figure 3 for the posterior predictive checks, showing that only the CHOCO model managed to capture the bimodal distribution of data.

### *Effect of Sex*

The ZOIB model suggested that women had higher mean (*mu* parameter) scores of reality beliefs ($\Delta_{\text{Female}} = 0.20$, 95% $CI$ $[0.03, 0.37]$, $pd = 98.77\%$), less extreme values (*zoi* parameter; STATS) but more ones relative to zeros (*coi* parameter; STATS).

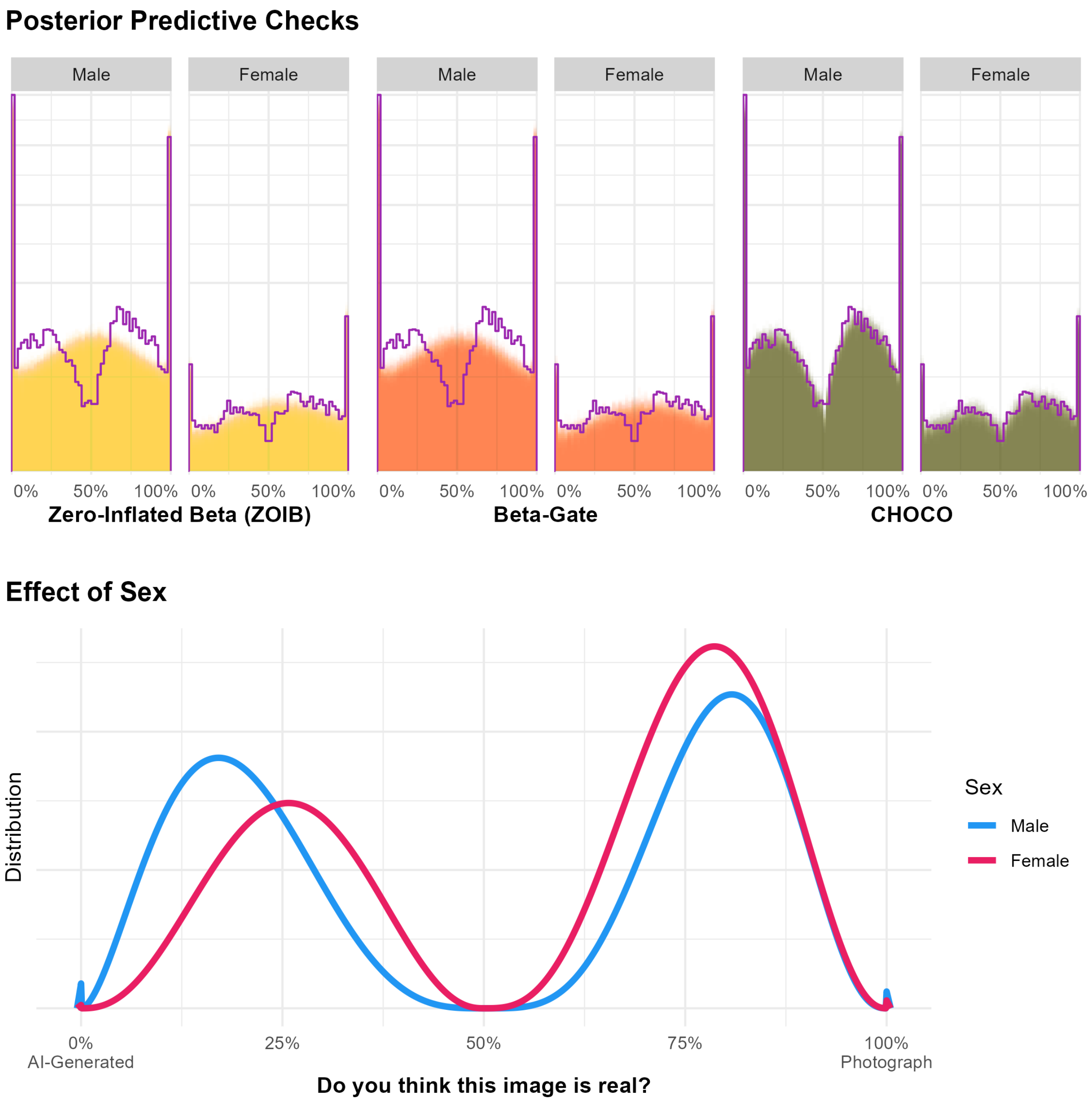
The Beta-Gate model similarly suggested that women had higher mean (*mu* parameter) scores of reality beliefs (STATS), less extreme values (*pex* parameter; STATS), and a greater tendency to answer one relative to zero (*bex* parameter; STATS).

The CHOCO model shows that women had a higher probability *p* of judging faces as real (STATS), but are not more confident when doing so (*confright* parameter; STATS). However, they were less confident when answering that an image was AI-generated (*confleft* parameter; STATS). There were also less likely to produce extreme answers (*pex* parameter; STATS), but no strong evidence supporting a directional bias was observed (*bex* parameter; STATS).

Across all these models, no effect of Sex on the precision parameter was observed.

**Figure 3**

*TODO.*



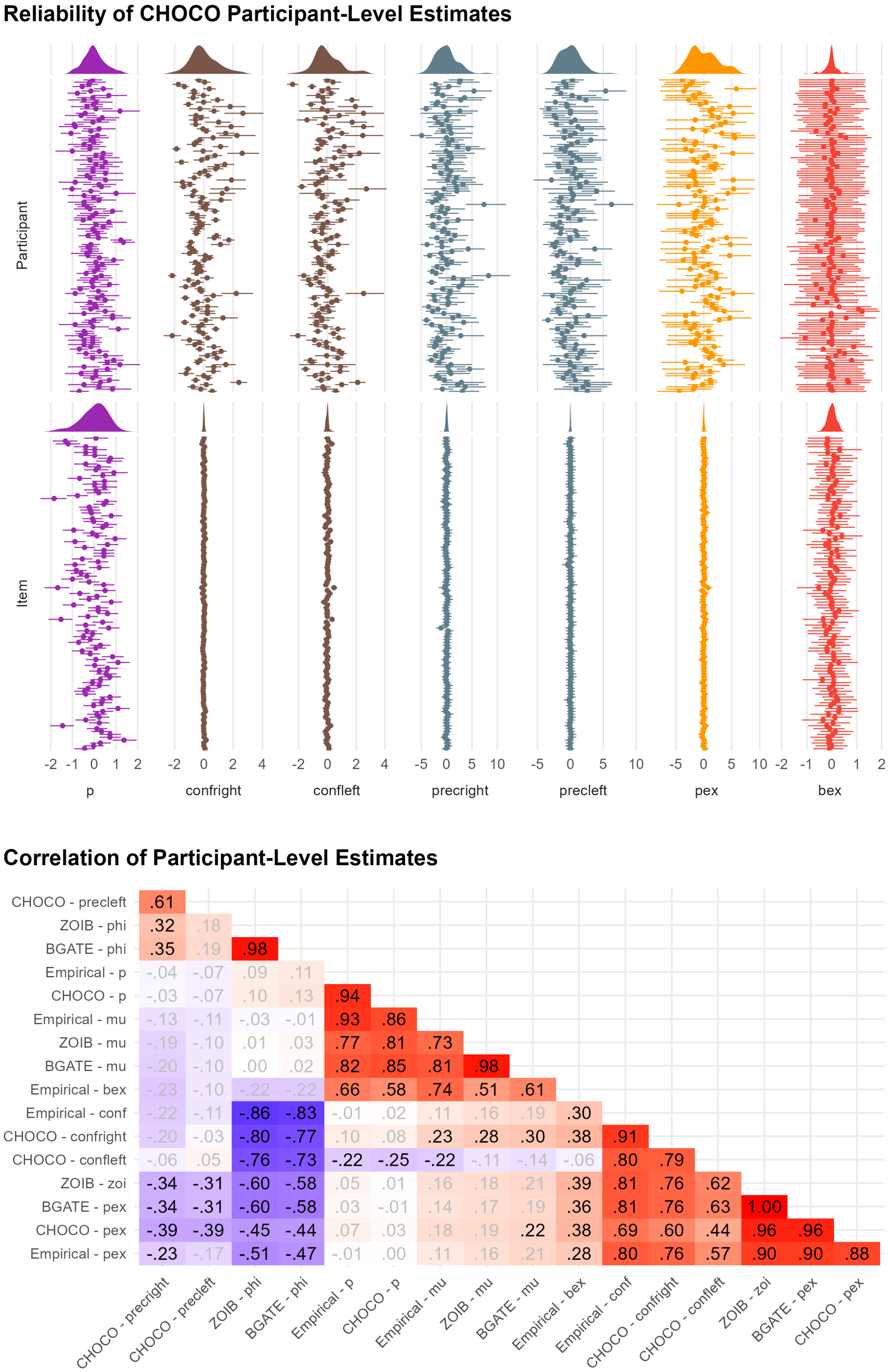
### *Individual-Level Parameters*

The ZOIB model estimated reliable variability in the participant's *phi* parameter (D-vour = 0.88) and *zoi* parameter (D-vour = 0.85), as well as in the *mu* parameter related to individual items (D-vour = 0.82). Moderate reliability was also observed for items in the *coi* parameter (D-vour = 0.71) and for participants in the *mu* parameter (D-vour = 0.69). The Beta-Gate model yielded similar results: a high reliability of participant's *phi* parameter (D-vour = 0.88), *pex* parameter (D-vour = 0.85). The *mu* parameter's variability was reliably captured for items (D-vour = 0.85) and moderately for participants (D-vour = 0.72).

The CHOCO model yielded reliable estimates for all pa-

**Figure 4**

*TODO.*



Reliability of CHOCO Participant-Level Estimates

Correlation of Participant-Level Estimates rameters except *bex* for participants (*confright* D-vour = 0.94, *confleft* D-vour = 0.91, *pex* D-vour = 0.79, *p* D-vour = 0.73, *precright* D-vour = 0.77, *precleft* D-vour = 0.67). Item's variability was primarily reflected through the *p* parameter (D-vour = 0.86).

Finally, the empirical average correlated the most strongly with CHOCO's *p* (r = .86), rather than ZOIB's *mu* (r = .77) or Beta-Gate's *mu* (r = .82). The empirical overall confidence was the strongest correlated with CHOCO's *confright* (r = .91), followed by ZOIB's *phi* (r = -.86), Beta-Gate's phi (r = -.83), and other CHOCO's parameters. The empirical proportion of "right" answer *p* correlated the strongest with CHOCO's *p* (r = 0.94), followed by Beta-Gate's *mu* (r = .82) and ZOIB's *mu* (r = .77). The empirical *pex* correlated the strongest with ZOIB's *zoi* (r = .90), and Beta-Gate's *pex* (r = .90), and CHOCO's *pex* (r = .88). Of note that the parameters of Beta-Gate and ZOIB correlate almost perfectly, underlining their empirical similarity despite a different parametrization.

## Discussion

Study 1 revealed that the Choice-Confidence (CHOCO) model was a much better fit for bimodal bounded data, compared to other alternatives like the Zero-and-One Inflated Beta (ZOIB) and Beta-Gate (Ordered Beta) models. It also allowed for a deeper understanding through its interpretable parameters, offering insights into possibly distinct cognitive mechanisms, such as the probability of answering real vs. fake, and the associated confidence in these two choices. This was illustrated by modelling the effect of sex on all the CHOCO parameters.

Note that the observed gender differences are primarily presented as a proof-of-principle, to showcase the model's ability to capture group-level effects and to provide deeper insights compared to other models. However, given that they were based on different items (female and male faces), these differences might just be a reflection of stimuli characteristics rather than true sex dymorphism in the formation of reality beliefs.

Finally, we also show that the CHOCO model was able to capture reliable and interpretable individual-level parameters, supporting its value to measure inter-individual differences. An interesting dissociation emerged between participant- and item-level variability: the latter seemed mostly to be represented in the *p* parameter, while participants reliably varied in most of the components (aside from *bex*). This could be interpreted as that external item characteristics primarily influence the probability of being judged as real vs. fake, while the expressed confidence is first and foremost an individual characteristic.

# Study 2

The explosion of accessibility of state-of-the-art AI tools has made it effortless to generate realistic images, including human faces that are often indistinguishable from real ones (Bozkir et al., 2024; Miller et al., 2023; Nightingale & Farid, 2022). These synthetic visuals are flooding the cyberspace across various domains, such as art, advertising and entertainment, to education and information. This technological advancement carries an important potential for misuse, such as in disinformation campaigns, scams (e.g., AI-bots, identity theft), and abuse. The democratization of such technology raises pressing concerns about the value of authenticity and the potential erosion of media trust in our increasingly *post-truth* society.

In this evolving landscape, understanding the cognitive mechanisms that underpin our judgments of reality becomes paramount. Despite the increasing prevalence of digitally altered or AI-generated content, humans still rely on certain heuristics to assess authenticity. One such heuristic might be facial attractiveness. Attractiveness appraisals are known to be automatic and unconscious (Hou et al., 2023; Hung et al., 2016; Luo et al., 2019), and carry strong real-life consequences, as demonstrated by the large body of literature on the "beauty premium" (Gulati et al., 2024; Kukkonen et al., 2024; Little, 2021; Pandey & Zayas, 2021)[3]

**TODO:** However, it is important to allow for the possibility of a U-shaped relationship. For instance, Peng et al. (2020) reported that both attractive and unattractive people sold more products (underlining a "plainness penalty").

Although its role as a potential modulator of reality beliefs remains underexplored, Makowski et al. (2025) found significant associations between participants' realness ratings and facial salience. Female participants judged both highly attractive and highly unattractive faces as more likely to be real, whereas for male participants, this effect was limited to highly attractive faces. These findings offer a complementary perspective to those of Miller et al. (2023), who reported that AI-generated faces tend to be rated as more attractive, and that participants used attractiveness as a distinguishing cue between real and AI-generated faces. Together, these results highlight a possible bidirectional and context-dependent role of attractiveness in shaping reality judgments.

In this study, we will apply the CHOCO model to the data of MAKOWSKI, and see whether different analysis approaches lead to different conclusions. We will then try to replicate these findings in a new sample of participants.

[3] As an eloquent example, Monk Jr et al. (2021) reported in a large representative US sample that the magnitude of earnings disparities among white women along the perceived attractiveness continuum exceeds in magnitude the canonical black-white race gap.

## Methods

### *Participants*

The first sample includes the same participants as study 1 (see above).

**TODO** for sample 2.

The final sample included 189 participants (Mean age = 28.4, SD = 14.0, range: [18, 69]; Sex: 76.2% females).

### *Procedure*

For the second sample, the procedure was relatively similar with a few key differences. Most importantly, subjective ratings were collected using a 7-point Likert scale ranging from 0 to 6, rather than a visual analog scale, which included a clear midpoint. Additionally, the familiarity rating was removed from the set of facial evaluations. In the second phase, realness judgments were also recorded on a 7-point Likert scale, this time ranging from -3 to +3, again with a defined midpoint, replacing the visual analog format used in the first study. For this scale, the endpoints were labeled 'AI-generated' and 'Photograph', replacing the 'Fake' and 'Real' anchors used in Study 1.

There were also modifications to the instructions provided to participants. In Phase 1, an additional prompt was introduced, stating that the study was conducted in partnership with a young AI startup. This was intended to strengthen participants' belief that the faces they were shown could plausibly be either real or AI-generated. Additionally, participants were explicitly asked to respond based on their immediate, first impressions—an instruction not included in the first study.

In Phase 2, participants were not informed about the proportion of AI-generated versus real faces encountered during Phase 1. This omission was deliberate, aimed at reducing demand characteristics and preventing participants from mechanically selecting half the faces as real and half as AI-generated. Instead, they were encouraged to base their responses solely on their level of confidence regarding the origin of each image.

### *Data Analysis*

To compare the benefits of CHOCO models to a "traditional" analytic approach, we started by fitting a frequentist linear mixed model to predict reality beliefs with the formula $Real \sim Sex/poly(Attractive, 2) + (poly(Attractive, 2)|Participant) + (1|Item)$. The 2nd degree raw polynomial term was included to allow for potentially non-linear relationships (note that the first and second degree effects of *raw* polynomials can be interpreted independently as the linear part and the "curvy" part of the relationship). For the CHOCO model, mildly informative and effect-agnostic (i.e., centered at zero) priors were used. The same formula was used for all parameters, except that (based on the reliability analysis of Study 1), items were only included as random effects for $p$, and participants were not included for $bex$.

**TODO**. For sample 2, the analysis was based on that of Sample 1. The main differences are 1) the inclusion of the "Condition" (whether the picture was presented as "Real" of "Fake" in the first phase of the experiment) as an additional predictor (it was entered as the only random slope for all participant random effects); 2) the inclusion of an additional parameters, *pmid*, modelling the probability of answering the middle-point of the Likert scale (representing an "undecided" option).

The same analysis was applied for Beauty.

## Results

As the complete model parameters tables are available at https://github.com/RealityBending/FictionChoco, we will focus on reporting noteworthy findings below.

### *Sample 1*

In sample 1, the traditional approach suggested a significant linear relationship between the mean level of "Reality" and attractiveness ($\beta_{poly1} = 3.42,\ 95\%\ CI[2.50, 4.34], p < .001\%$) for males only. The CHOCO model revealed that attractiveness had a significantly positive linear relationship with the probability $p$ of judging faces as real (STATS), but a quadratic relationship with the confidence in real judgments (*confright*; STATS). Attractiveness was also associated with less confidence in fake judgments (*confleft*; STATS), a quadratic relationship with left precision (*precleft*; STATS), and related linearly with a stronger bias towards extreme Real responses (*bex*; STATS). These effects were present for males only, and no relationship was found for females. The reliability of the effect of attractiveness (the random slopes) was very low for all parameters (D-vour < 0.01). The effect of Beauty was the same as of attractiveness.

### *Sample 2*

In Sample 2, the traditional approach suggested a significant linear relationship between the mean level of "Reality" and attractiveness for males (STATS) and women (STATS), with no effect of the Condition. The CHOCO model revealed that attractiveness had a significantly positive linear relationship with the probability $p$ of judging faces as real for males (STATS) and females (STATS). It also had a linear relationship with the confidence in real judgments (*confright*) for males (STATS) and females (STATS), as well as a significant quadratic curvature for males only (STATS). Attractiveness also linearly decreased the confidence in fake judgments only for males (*confleft*; STATS), but increased the probability of an undecided response (*pmid*) in women (STATS), and displayed in females only a quadratic relationship with the amount of extreme answers (*pex*; STATS) and its bias in favour of right-side extreme response (*bex*; STATS). No consistent and reliable effect of condition was found.

The effect of Beauty was…

### Discussion

The fact that the effect was less obvious in women in the sample 1 might be explained with power (less relevant items were included for women).

## General Discussion

This study introduces a new statistical model to analyze bimodal data, which can be observed for instance when using subjective rating scales in which the two sides correspond to a different "choice" (e.g., "false" vs. "true", "real" vs "fake"). The Choice-Confidence (CHOCO) model conceptualizes the data as a mixture of two distributions, one for each choice, and models the probability of choosing one or the other as well as the degree of confidence in each choice. We validated this model on real-life data by appling it to the reality beliefs of Makowski et al. (2025), and showed that it was able to capture the bimodal nature of the data and provides a more accurate representation of the underlying processes than other alternative approaches, including Zero-or-One inflated Beta (ZOIB) models.

Little difference between Beauty and Attractiveness, aside that Beauty revealed more effects in women. Why?

We demonstrated the flexibility of the CHOCO model, being usable in analog scales as well as discrete ratings with a mid-point.

Likert-scales (with a limited number of discrete options) are commonly used in psychology, and often modeled as continuous. While this fact has spurred its own polarized debate (REFS), our position is that the model used should ideally reflect the data *generating* mechanism rather than necessarily focus on the *observed* data[4] - although the former cannot always be easily inferred or known. For the data of Sample 2, we held the assumption that the underlying latent distribution was CHOCO distributed (as there is no reason to assume it would be different from Sample 1), treated as continuous CHOCO-distributed data, despite the fact that it was collected on a 7-point Likert scale. The motivation was to see if the model could extrapolate and reliably estimate all the parameters based on a limited number of data points but means in our case that, effectively, each Beta distribution was extrapolated from the frequency of only two unique values (or 3 for the no-extreme version). While this is far from ideal, the CHOCO model still provided a better fit to the data than alternative approaches. **TODO** However, it might have lead to a lesser ability to detect finer grain changes, including U-shaped relationships. Having 7-points (or 6 with no mid-value) is the strict minimum to use the CHOCO model.

Future studies should assess psychometric property liek the minimum amount of data required to estimate all parameters reliably.

Future studies should assess the ability of CHOCO models to also fit unimodal distributions (so it can be used more generally), as well as the benefits of more parsimonious parametrization (for instance, a unique precision parameter controlling both the left and right sides, or a linked confidence parameter e.g. where *confleft* is expressed as a function of *confright*).

## Data Availability

Data, code and everything is available at https://github.com/RealityBending/FictionChoco. The CHOCO model is implemented in the *cogmod* R package (https://github.com/DominiqueMakowski/cogmod).

## Acknowledgements

We would like to thank the dissertation students from the University of Sussex for their help in data collection. DM would also like to thank … for the motivation provided to write this paper.

## References

Azevedo, R., Tucciarelli, R., De Beukelaer, S., Ambroziak, K., Jones, I., & Tsakiris, M. (2020). *A body of evidence:'feeling in seeing'predicts realness judgments for photojournalistic images.*

Bozkir, E., Riedmiller, C., Skodras, A. N., Kasneci, G., & Kasneci, E. (2024). Can you tell real from fake face images? Perception of computer-generated faces by humans. *ACM Transactions on Applied Perception*, *22*(2), 1–23.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28.

Chen, J. M., Norman, J. B., & Nam, Y. (2021). Broadening the stimulus set: Introducing the american multiracial faces database. *Behavior Research Methods*, *53*, 371–389.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Gulati, A., Martínez-Garcia, M., Fernández, D., Lozano, M. A., Lepri, B., & Oliver, N. (2024). What is beautiful is still good: The attractiveness halo effect in the era of beauty filters. *Royal Society Open Science*, *11*(11), 240882.

[4] Ideally, the data recording method should be aligned with the assumed generating mechanism, which is an experiment design issue rather than a statistical one.

Hou, X., Shang, J., & Tong, S. (2023). Neural mechanisms of the conscious and subliminal processing of facial attractiveness. *Brain Sciences*, *13*(6), 855.

Hung, S.-M., Nieh, C.-H., & Hsieh, P.-J. (2016). Unconscious processing of facial attractiveness: Invisible attractive faces orient visual attention. *Scientific Reports*, *6*(1), 37117.

Kubinec, R. (2023). Ordered beta regression: A parsimonious, well-fitting model for continuous data with lower and upper bounds. *Political Analysis*, *31*(4), 519–536.

Kukkonen, I., Pajunen, T., Sarpila, O., & Åberg, E. (2024). Is beauty-based inequality gendered? A systematic review of gender differences in socioeconomic outcomes of physical attractiveness in labor markets. *European Societies*, *26*(1), 117–148.

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369.

Little, A. C. (2021). Facial attractiveness. *Encyclopedia of Evolutionary Psychological Science*, 2887–2891.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An r package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60).

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Wiernik, B. M., Bacher, E., Thériault, R., & Makowski, D. (2022). Easystats: Framework for easy statistical modeling, visualization, and reporting. *CRAN*. https://doi.org/10.32614/CRAN.package.easystats

Luo, Q., Rossion, B., & Dzhelyova, M. (2019). A robust implicit measure of facial attractiveness discrimination. *Social Cognitive and Affective Neuroscience*, *14*(7), 737–746.

Makowski, D. (2018). *Cognitive neuropsychology of implicit emotion regulation through fictional reappraisal* [PhD thesis]. Sorbonne Paris Cité.

Makowski, D., Ben-Shachar, M. S., Chen, S. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in Psychology*, *10*, 2767.

Makowski, D., Sperduti, M., Pelletier, J., Blondé, P., La Corte, V., Arcangeli, M., Zalla, T., Lemaire, S., Dokic, J., Nicolas, S., et al. (2019). Phenomenal, bodily and brain correlates of fictional reappraisal as an implicit emotion regulation strategy. *Cognitive, Affective, & Behavioral Neuroscience*, *19*, 877–897.

Makowski, D., Te, A. S., Neves, A., Kirk, S., Liang, N. Z., Mavros, P., & Chen, S. A. (2025). Too beautiful to be fake: Attractive faces are less likely to be judged as artificially generated. *Acta Psychologica*, *252*, 104670.

Makowski, D., & Waggoner, P. D. (2023). Where are we going with statistical computing? From mathematical statistics to collaborative data science. In *Mathematics* (8; Vol. 11, p. 1821). MDPI.

Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A., Krumhuber, E. G., & Dawel, A. (2023). AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychological Science*, *34*(12), 1390–1403.

Monk Jr, E. P., Esposito, M. H., & Lee, H. (2021). Beholding inequality: Race, gender, and returns to physical attractiveness in the united states. *American Journal of Sociology*, *127*(1), 194–241.

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, *119*(8), e2120481119.

Ospina, R., & Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, *56*(6), 1609–1623.

Pandey, G., & Zayas, V. (2021). What is a face worth? Facial attractiveness biases experience-based monetary decision-making. *British Journal of Psychology*, *112*(4), 934–963.

Peng, L., Cui, G., Chung, Y., & Zheng, W. (2020). The faces of success: Beauty and ugliness premiums in e-commerce platforms. *Journal of Marketing*, *84*(4), 67–85.

Rouder, J. N., & Mehrvarz, M. (2024). Hierarchical-model insights for planning and interpreting individual-difference studies of cognitive abilities. *Current Directions in Psychological Science*, *33*(2), 128–135.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432.

Williams, D. R., Mulder, J., Rouder, J. N., & Rast, P. (2021). Beneath the surface: Unearthing within-person variability and mean relations with bayesian mixed models. *Psychological Methods*, *26*(1), 74.