Reply to:
Comment on "Decisions under Uncertainty as Bayesian Inference on Choice Options" (initial submission)

Ferdinand M. Vieider

22 January 2026

Preliminary statement

Below, I summarize the key points raised in a comment by Kemel and Richard (hereafter KR) on my paper published in Management Science (Vieider, 2024). The present document responds to the version of the comment initially submitted by KR. For transparency, I briefly summarize the sequence of events as known to me:

- In July 2025, Kemel and Richard informed me that they had submitted a comment on my paper to Management Science.
- As an empirical illustration of my model, my paper replicates the "mirror results" reported by Oprea (2024) in a binary choice setting.
- KR's comment, inter alia, reiterates concerns previously expressed by George Wu, Uri Simonsohn, and coauthors regarding that result.
- Several months later, I learned that the comment was under review at Management Science.
- I was not initially included in the review process for the comment.
- I raised this issue with the editor in chief, Christoph Loch.
- After multiple email exchanges, Prof. Loch informed me that George Wu was the handling editor.
- Prof. Loch subsequently invited me to submit a report directly to him.
- I was not informed of the identity of the treating associate editor.
- I submitted a report under the understanding that I would be kept informed of subsequent developments.
- I have not received further communication from Management Science regarding the comment.
- I have since learned indirectly that a revision of the comment has been invited.

Given this situation, I make public responses to all versions of the comment known to me. This is intended both to document my position and to contribute to open scientific discussion.

Reply to comment (version dating to July 2025)

Summary. Kemel and Richard (henceforth *KR*) propose a comment on Vieider (2024). Their key claim is that the model parameters — and in particular, the likelihood-discriminability parameter $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$, the outcome-discriminability parameter $\alpha \triangleq \frac{\sigma_o^2}{\sigma_o^2+\nu^2}$, and the contribution of the prior, $\delta \triangleq \left(\frac{\psi}{1-\psi}\right)^{1-\gamma}$, are not uniquely identified. Based on this alleged lack of identification, they revisit several empirical results: 1) they use the supposed non-identifiability to propose an amendment of the model, and claim that this changes the conclusion for the experiment pitching risk against mirrors; 2) they use the alleged absence of identifiability to claim that my model makes no quantitative predictions on correlations between errors and likelihood-insensitivity estimated from PT, and revisit this based on what they call a "meta-analysis". Finally, 3) they include some comments on utility curvature under risk and ambiguity.

Assessment. KR's claim of lack of identifiability of the model parameters is difficult to reconcile with the fact that Vieider (2024) explicitly introduces parameter restrictions with the aim of guaranteeing identifiability. The parameter definitions, briefly repeated in the summary above, make clear that the assumption of common coding noise creates a common scale for the two discriminability parameters α and γ. These parameters are fully determined by the signal-to-noise ratios σ_o/ν and σ_p/ν, and are thus measured on the scale of the common coding noise ν. Using simulations, it is straightforward to establish a basic fact: parameter recoverability is, in fact, excellent. Against this background, it is difficult not to be puzzled by the claim of non-identifiability.

Given that the parameters are identified, the claims concerning mirrors versus risk do not follow. KR nonetheless advance additional arguments in this context. One claim is that their estimations based on a restricted version of my BIM lead to "different conclusions" (p. 2). The conclusions they report, however, largely reiterate statements already contained in my paper. Another claim is that "estimates take arbitrary values driven by priors, as these parameters cannot be estimated on the basis of the likelihood function alone" (p. 3). This claim is difficult to sustain: re-estimating the model under a wide range of prior specifications, including improper priors so that inference is governed by the likelihood alone, produces no meaningful changes in the estimated parameters. It therefore remains unclear on what basis the claim that results are driven by priors is made.

The claims regarding utility differences under risk and ambiguity largely miss the purpose of the original exercise. While a more detailed investigation of this issue could indeed be of interest, KR do not pursue such an investigation. Instead, they combine data obtained from a bisection procedure—which is not described by my model—with the original data, and then search for additional tests under which the resulting estimates fail to reach conventional significance thresholds. Several aspects of this analysis raise concerns, including reliance on well-known statistical pitfalls. Most importantly, however, this discussion overlooks the broader point: the specific violation of probability–outcome separability examined in Vieider (2024) is only one among several such violations documented in the literature, all of which are explicitly shown to be predicted by the BIM. These broader implications are not addressed.

Finally, KR revisit correlations between likelihood-discriminability $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$ and the

standard deviation of response errors, $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$. This argument builds on the claim that the BIM does not generate quantitative predictions for such correlations. As shown below, this claim is incorrect: the model does generate clear quantitative implications once its equations are taken into account. The analysis presented by KR, including what they describe as a "meta-analysis", does not engage with these implications and reflects well-known but unresolved issues concerning the econometric recoverability of prospect theory parameters in random-utility settings—issues that the BIM is designed to address.

In sum, the central claims that the model parameters are not identified, that the results are driven by priors, and that the BIM makes no quantitative predictions regarding correlations between noise and discriminability are objectively false. Below, I show in detail how a careful examination of the model and the data establishes this beyond reasonable doubt. Some of the empirical questions raised—particularly regarding utility differences and correlations—could in principle warrant further investigation. At present, however, the analyses presented by KR do not provide new evidence bearing on these issues.

1 Identifiability and number of parameters

The central claim advanced by KR concerns the identifiability of the BIM parameters. A closely related secondary claim is that the BIM contains too many parameters. In this section, I first show that the claim of non-identifiability is factually incorrect. I then argue that simple parameter counting and comparisons to prospect theory (PT) are not meaningful in this context, because *the parameters in the two models are of a fundamentally different nature.*

Parameter identifiability. KR state that "these parameters [α and γ] are only unique up to a constant factor", and that "it is impossible to separate α from γ" (p. 3). These statements are difficult to reconcile with two features of the model that are explicitly discussed in Vieider (2024). First, the paper motivates modelling coding noise as common to log-odds and log cost-benefits precisely in order to guarantee identifiability ("the coding noise is modelled as being common to the two dimensions because different coding noises would not be separately identifiable from direct tradeoffs", p. 9018). Second, the definitions $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$ and $\alpha \triangleq \frac{\sigma_o^2}{\sigma_o^2+\nu^2}$ place both parameters on the common scale of the coding noise ν. Inspection of these definitions already suffices to establish that the parameters are identified. The assumption of common coding noise is isomorphic to the correlation structure used by Natenzon (2019) to guarantee identifiability in the multinomial probit model.

To address any remaining doubts, identifiability can be assessed directly by examining the econometric recoverability of parameters from simulated data. I therefore study the recoverability of $\alpha \triangleq \frac{\sigma_o^2}{\sigma_o^2+\nu^2}$, $\gamma \triangleq \frac{\sigma_p^2}{\sigma_p^2+\nu^2}$, $\delta \triangleq \xi^{1-\gamma}$, and the coding noise ν using simulations. I simulate 100 subjects with parameters drawn independently as follows:

$$\nu \sim \left|\mathcal{N}(0.6, 0.3)\right|, \quad \sigma_p \sim \left|\mathcal{N}(0.7, 0.35)\right|, \quad \sigma_o \sim \left|\mathcal{N}(0.9, 0.4)\right|, \quad \xi \sim \left|\mathcal{N}(0.7, 0.3)\right|.$$

The resulting distributions feature both reasonably high average signal-to-noise ratios and substantial variation in the parameters whose identifiability is questioned, making

the recovery exercise informative. I estimate the BIM using non-informative hyperpriors on the population-level parameters, specified as normal distributions with mean zero and standard deviation ten on the log scale. These priors assign 95% of their mass to the interval between 0 and $\exp(1.96 \cdot 10) \approx 3.25 \times 10^8$, which for practical purposes is effectively unbounded. The econometric code otherwise follows Vieider (2024). Each simulated subject completes 300 binary choices with orthogonal variation in log-odds and log cost-benefits.

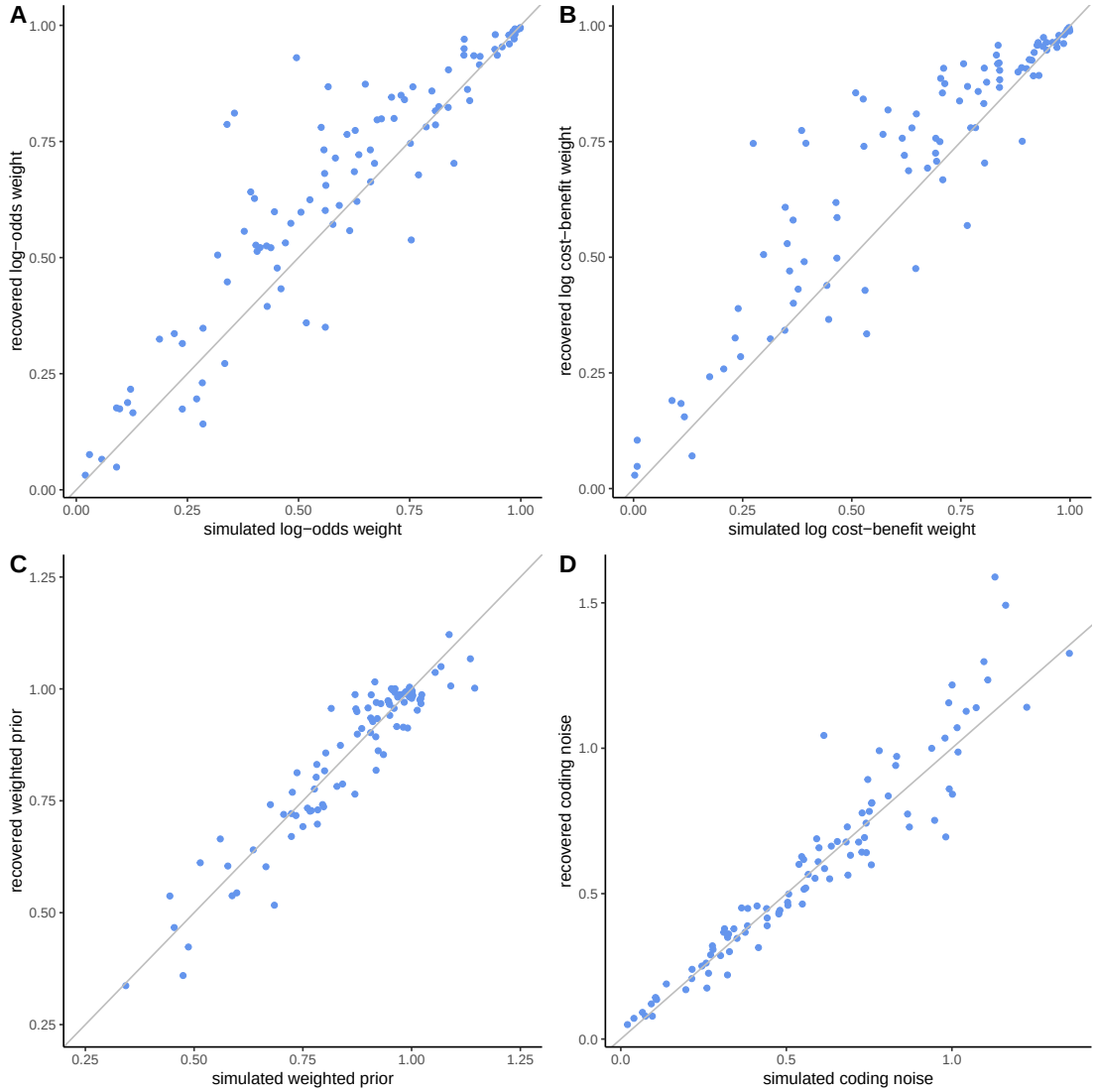

Figure 1: Scatter plots of simulated versus recovered parameters

The figure shows correlations between simulated and recovered parameters: likelihood-discriminability γ (panel A), outcome-discriminability α (panel B), the weighted prior mean δ (panel C), and coding noise ν (panel D). Gray lines indicate equality between simulated and recovered values.

Contrary to KR's claim that "it is impossible to separate α from γ", parameter recoverability is excellent. Figure 1 shows strong correspondence between simulated and recovered parameters, with all Spearman correlations exceeding 0.9. The recovered parameters closely match the simulated magnitudes, providing direct evidence against the claim of non-identifiability. These correlations are near the upper bound achievable given noisy choice data. It therefore remains unclear on what basis the claim of non-identifiability

is made.

Number of parameters. KR also raise concerns about the number of parameters in the BIM. This criticism rests on treating the BIM parameters as analogous to those of standard descriptive models such as prospect theory. Such a comparison is not appropriate in this context, because the parameters in the BIM are explicitly *generative* rather than descriptive, as I emphasize repeatedly in my paper. Here, I revisit KR's claims and illustrate the difference between parameters in the different model classes.

KR state that the model contains five parameters, while Vieider (2024) estimates four. Regardless of the exact count, this discussion misses the central point of the model. As emphasized repeatedly in Vieider (2024), PT parameters are introduced to fit choice data, whereas BIM parameters correspond to latent components of a cognitive inference process that causally generates choice. In dynamic environments, separate priors are therefore essential for prediction, as illustrated by Bouchouicha et al. (2025) using a six-parameter version of the model of Khaw et al. (2021).

To illustrate the contrast, consider parameterizations of PT for choices over gains. Depending on modelling choices, PT can be estimated using three parameters (utility curvature, probability weighting, and noise), four parameters (adding elevation of the weighting function), five parameters (e.g., using an expo-power utility function), or more. In practice, parameter counts are often chosen post hoc to improve fit, with parameters being largely independent and only weakly constrained. Recent work by Balcombe and Fraser (2025), for example, estimates a 17-parameter specification to fit choices over gains, losses, and mixed gambles.

KR implicitly treat BIM parameters as belonging to this descriptive class. However, given the generative nature of the BIM, the parameters have a different interpretation. They correspond to pre-choice cognitive quantities that govern inference about latent choice primitives. A fully specified version of the model could in principle include additional parameters, such as dimension-specific coding noise for outcomes and probabilities. This does not imply that all such parameters must be recoverable from binary choice data alone. The four parameters estimated in the journal version are explicitly chosen to be econometrically identifiable, as demonstrated above.

The generative interpretation of BIM parameters is further supported by independent evidence from neuroscience. A substantial literature identifies noise signatures in neural activation and links them to subsequent behavior (see, e.g., Van Bergen et al., 2015; Barretto-García et al., 2023). Using neuroimaging methods that separately manipulate probabilities and outcomes, de Hollander et al. (2024) are able to identify prior means in the model of Khaw et al. (2021) on a trial-by-trial basis.

Several recent papers further illustrate the implications of this generative interpretation. In Oprea and Vieider (2024), probability dependence in choice is established in a model-free way. Once subjects are forced to take representative samples from fully described choice options, however, probability dependence in risky choice vanishes. While such sampling is redundant from the perspective of standard descriptive models, the sampling-based BIM predicts a reduction in coding noise and thus a collapse of probability dependence—precisely what is observed.

To further illustrate this point, consider how one would analyze the experimental results

in Oprea and Vieider (2024) using PT, which makes no predictions in this context. If one were to apply the same four-parameter specification estimated in the control treatment to the sampling treatment, the resulting fit would be extremely poor. Put differently, the control treatment has very little predictive value for behaviour in the sampling treatment. One could, of course, re-estimate a new set of four parameters ex post to accommodate the observed behavioural differences. Apart from being arbitrary, this approach would effectively introduce eight free parameters across the two treatments.

By contrast, the BIM relies on the same underlying parameters in both treatments. The only change is that coding noise is endogenously updated through sampling, leading to a substantial reduction in ν. This illustrates why a discussion focused solely on the number of parameters misses the substantive point: what matters is not parameter count per se, but whether a model generates cross-treatment predictions based on a coherent underlying mechanism.

Similarly, Oprea and Vieider (2025) orthogonally manipulate coding noise in probability and outcome dimensions. Without estimating prior means, they show that changes in coding noise generate differential predictions via their effect on the weight placed on priors. Large behavioral changes, including reversals in probability dependence, are predicted and observed without introducing additional parameters. These results do not rely on particular assumptions about priors, which are treated as latent and invariant across conditions; instead, differences across conditions arise from experimentally induced changes in coding noise.

Conclusion on identifiability. KR's claim regarding the lack of identifiability of the BIM parameters is incorrect. Since several of their subsequent arguments build on this claim, its failure has direct implications for their empirical conclusions. I now turn to these empirical points.

2 Risk versus mirrors

I now turn to the claims KR make regarding the experiment comparing risk to mirrors. Since their central claim concerning lack of identifiability does not hold, the implications they draw from it do not follow. KR nonetheless advance additional arguments, in particular that their alternative specification changes the conclusions, and that the estimates reported in Vieider (2024) depend on prior assumptions. I address these points in turn.

No new evidence. KR claim that their modified version of the BIM leads to different conclusions. A formulation based on the ratio γ/α is indeed behaviorally meaningful, and is explicitly discussed in Vieider (2024) (see equation 13 and surrounding text). Bouchouicha et al. (2024) similarly adopt this formulation to illustrate the contrast between binary choice and certainty equivalents, and such a specification forms the key equation on which the predictions of Oprea and Vieider (2025) are based.

Contrary to KR's claim, however, the key results reported in Vieider (2024) are based entirely on non-parametric tests of choice proportions. The finding that log-odds weights γ are larger for mirrors than for risk—presented by KR as altering the conclusions—is already documented in Figure 5a of the original paper. No conclusions in Vieider (2024)

rely on model-based estimation for this comparison. The paper explicitly notes that the pattern in Figure 5a may indicate "an effect of risk *in addition to* a complexity effect" (p. 9025, emphasis in original). In this sense, the results reported by KR closely mirror statements already contained in Vieider (2024).

More broadly, the purpose of the mirror treatment is not to provide a direct test of the BIM, but to examine whether increased complexity alone can generate choice patterns that resemble what are commonly interpreted as risk attitudes. The aim is to replicate the results of Oprea (2024) using binary choice rather than choice lists, and to contribute to a broader literature on noisy cognition and efficient adaptation (see, e.g., Steiner and Stewart, 2016; Herold and Netzer, 2023; Netzer et al., 2024), as explicitly discussed in Vieider (2024).

This question is currently the subject of active debate. For example, the interpretation proposed by Banki, Simonsohn, Walatka and Wu (2025), which relies on switching behavior within choice lists, does not apply to binary choice data. In particular, the finding of expected-value maximization for very small probabilities is difficult to reconcile with such an explanation. These considerations highlight that the interpretation of mirror-risk differences is non-trivial and requires careful analysis.

Finally, it remains unclear what any observed differences between risk and mirrors should be taken to imply. Such differences could reflect stable preferences, or they could arise from differences in the normative content of value maximization under certainty versus *expected* value maximization under risk. These issues are well worth further investigation. However, the analyses presented by KR do not introduce new evidence that would help adjudicate between these interpretations.

Identification driven by priors. KR further claim that the parameter estimates reported in Vieider (2024) are driven primarily by restrictive priors. This claim can be assessed directly.

The priors used in Vieider (2024) are specified on population-level parameters and thus function as *hyperpriors*. They are chosen to be mildly informative in order to facilitate convergence of the estimation algorithm, which is standard practice in Bayesian hierarchical modeling (see, e.g., McElreath, 2016). Given the amount of data available, large prior effects on the estimated parameters would not be expected. Nonetheless, the statement by KR that "estimates take arbitrary values driven by priors, as these parameters cannot be estimated on the basis of the likelihood function alone" (p. 3) can be tested explicitly.

To do so, I estimate (i) the version of the BIM underlying the published results in Vieider (2024), and (ii) the same model using improper hyperpriors spanning the full support of the parameters. The latter corresponds to inference driven entirely by the likelihood and is therefore equivalent to a frequentist estimation.

Figure 2 shows that the parameter estimates obtained under informative and improper priors are almost perfectly correlated, with no systematic differences in magnitude. These results provide no support for the claim that the estimates are driven by prior assumptions. Instead, they indicate that inference is dominated by the likelihood, contrary to the assertion made by KR.

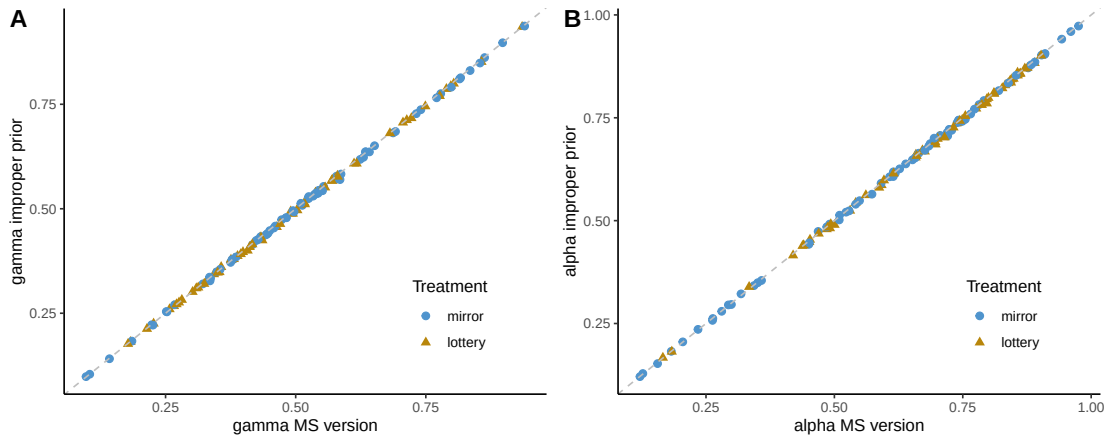

Figure 2: Scatter plots of parameters estimated under informative versus improper priors

Scatter plots of parameter estimates obtained using the priors reported in Vieider (2024) (x-axis) and estimates obtained under improper hyperpriors (y-axis). Panel A shows likelihood-discriminability γ, and Panel B outcome-discriminability α. Correlations are close to unity in both cases.

Conclusion on lotteries versus mirrors. KR further analyze what they describe as a "new version" of the BIM. However, this formulation is already presented in equation 13 of Vieider (2024) and discussed in the surrounding text. Moreover, as shown above, the fully specified model is identified. The differences between risk and mirrors documented by KR in their parametric analysis correspond closely to patterns already discussed in Vieider (2024). Finally, their claim that parameter estimates are driven by restrictive priors is not supported by the evidence once the issue is examined directly.

3 Utility under risk and ambiguity

The discussion of utility differences between risk and ambiguity is largely orthogonal to the main arguments of the comment. KR augment the analysis by incorporating data from Abdellaoui et al. (2011) and by proposing alternative statistical tests. In this section, I argue that these additions do not provide new insight into the question at hand and that several of the conclusions drawn rely on common statistical pitfalls. I conclude with more constructive remarks, since this question could in principle be resolved straightforwardly with appropriately designed new data.

Main point and statistical issues. The utility comparison in Vieider (2024) is not intended to provide a strong or definitive test of the BIM. First, it constitutes a relatively indirect prediction of the model. Second, the paper explicitly presents the results on risk versus ambiguity jointly with a broader body of existing evidence on violations of probability–outcome separability, which are predicted by the BIM but not by prospect theory. The illustrative force of this argument derives from the accumulation of prior empirical findings (Hogarth and Einhorn, 1990; Fehr-Duda et al., 2010; Bouchouicha and Vieider, 2017), a point that is stated explicitly at multiple places in the paper.

The data from Abdellaoui et al. (2011) used by KR are obtained using a bisection procedure. Apart from concerns regarding incentive compatibility, this method requires subjects to map binary choices onto underlying choice lists via non-trivial instructions. From a cognitive perspective, such mappings would themselves require explicit modelling. As a result, these data fall outside the scope of the BIM, which is a model of

binary choice. Consistent with this, the meta-analysis of Imai et al. (2025) documents systematically different prospect-theory parameter estimates for bisection data compared to those obtained from binary choice or certainty equivalents.[1]

It is therefore not surprising that some alternative tests yield non-significant results for the relatively small experiment reported in Vieider (2024). At most, the additional analyses reported by KR add two non-significant tests to two significant ones. Moreover, the tests they employ are less standard than those used in the original paper. In particular, Wald tests are known to perform poorly in small samples, and the use of predictive-performance tests to assess parameter differences is unconventional. In drawing strong conclusions from these results, the analysis also reflects two well-known statistical fallacies: (i) interpreting the difference between a significant and a non-significant result as itself statistically significant (Gelman and Stern, 2006), and (ii) treating the absence of evidence as evidence of absence.

Constructive suggestions. The question of whether utility differs between risk and ambiguity is both meaningful and empirically tractable. Addressing it convincingly would require a design that allows both acceptance and rejection of the null hypothesis. This would entail, at a minimum: (i) a sufficiently large sample size; (ii) a rich and well-calibrated set of stimuli; (iii) a measurement method that does not favor one theoretical prediction ex ante; and (iv) an explicit framework for testing equivalence rather than relying on non-significance. Such an approach could provide a principled resolution of the issue. The analyses presented by KR, however, do not meet these requirements and therefore do not advance the debate.

4 Correlations between noise and likelihood sensitivity

The discussion of correlations between likelihood sensitivity and stochastic choice noise is potentially the most substantive part of the comment by KR. However, the analysis presented there suffers from three distinct problems. First, the empirical aggregation labeled a "meta-analysis" lacks a clear definition of units and methods. Second, the interpretation of the observed correlations relies on the same conceptual assumptions that underlie earlier claims regarding identifiability and priors. Third, the analysis does not take into account well-known difficulties in recovering prospect-theory (PT) parameters from binary choice data using random utility models (RUMs), difficulties that are closely related to the mechanisms captured by the Bayesian inference model (BIM). I address these issues in turn.

Study selection and aggregation. The aggregation exercise referred to as a "meta-analysis" is not accompanied by a description of inclusion criteria, weighting schemes, or estimation methods. Several of the datasets included rely on bisection procedures, despite statements to the contrary in the comment. Such data are not directly compatible with the BIM, which is formulated for binary choice, and they are known to generate systematically different PT parameter estimates than binary choice or certainty equivalents (Imai et al., 2025). Moreover, it is unclear how the unit of analysis is defined.

[1]The BIM is formulated for binary choice. Vieider (2024) also applies it to certainty equivalents, based on the intuition that binary choices grouped in lists activate similar—though not identical—cognitive mechanisms. This intuition has since been formalized in extensions of the BIM to choice lists by Bouchouicha et al. (2024) (see also Khaw et al., 2023).

For example, losses are excluded even though they are included in Vieider (2024), while multi-experiment papers are treated inconsistently across sources. The absence of a documented aggregation method makes it difficult to assess the meaning of the reported correlations. Standard issues in meta-analysis—such as dependence across estimates, model heterogeneity, and publication bias—are not addressed (see, e.g., Borenstein et al., 2009; Brown et al., 2024).

A comprehensive meta-analysis of correlations between noise and likelihood sensitivity could in principle be informative. Existing surveys already document the universe of PT estimates based on certainty equivalents and binary choice (Imai et al., 2025; Bouchouicha et al., 2024). A defining feature of meta-analysis is that it aims to include all relevant evidence. Any such effort, however, would need to be grounded in a clear understanding of the quantitative predictions of the underlying model.

Conceptual issues. KR state that "it is impossible to predict the magnitude" of correlations between likelihood sensitivity $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$ and response noise $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$ within the BIM. This claim follows from treating noise as an additively separable disturbance, as in standard RUMs. In the BIM, by contrast, the relationship between coding noise and response variability is structurally determined by Bayesian inference. here in sequence, I discuss some of the key issues. The discussions follow the examination of stochastic choice structures by Vieider (2026).

For given prior variances σ_p and σ_o, response noise is a non-monotonic function of coding noise ν. As coding noise increases for fixed prior variances, decision-makers place increasing weight on the prior, which eventually reduces behavioural variability. This property is a direct implication of Bayesian inference and is shared by all models defining stochastic choice based on Bayesian inference, such as e.g. (Khaw et al., 2021). The intuition and formal derivation are discussed in detail by Ma et al. (2023). In Vieider (2026), I provide empirical tests and find strong support for Bayesian inference.

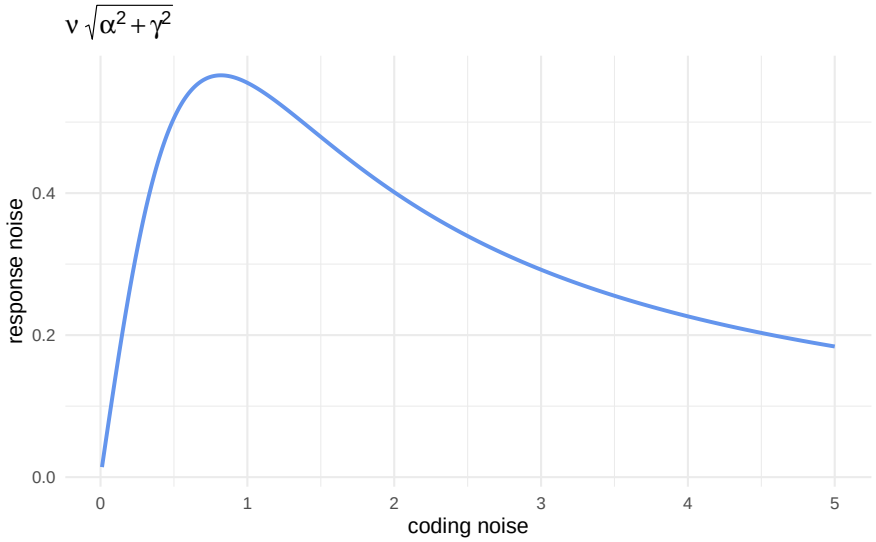

Figure 3: Response noise as a function of coding noise

The figure illustrates the non-monotonic relationship between response noise ω and coding noise ν in the BIM. This feature ensures stochastic monotonicity of choice behavior, in contrast to standard random utility models.

Figure 3 illustrates this relationship for $\sigma_p > \sigma_o$. Response noise initially increases with coding noise but eventually declines as choices converge toward the prior mean. This property, in turn, mean that the mapping between aggregate choice-generating parameters (such as α, γ, and δ) and stochastic choice proportions of the lottery is monotonic.

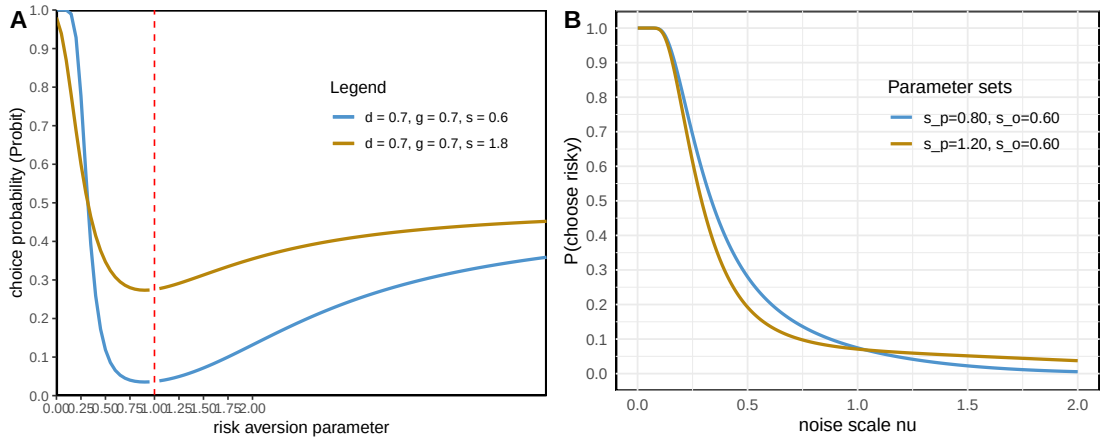

Figure 4: Stochastic monotonicity in RUMs and the BIM

Panel A illustrates stochastic non-monotonicity in a PT-based RUM as utility curvature increases. Panel B shows that choice behavior in the BIM remains monotonic as coding noise increases.

Figure 4 contrasts stochastic choice patterns in RUMs and the BIM. In RUMs, increasing risk aversion can generate non-monotonic choice probabilities, leading to identification problems (Apesteguia and Ballester, 2018). In the BIM, by contrast, choice probabilities vary monotonically with coding noise (Vieider, 2026).

Differences between binary choice and certainty equivalents naturally arise because binary choice typically exhibits a higher signal-to-noise ratio. This prediction is borne out empirically by Bouchouicha et al. (2024), who document systematically higher estimated likelihood sensitivity in binary choice than in certainty equivalents using non-parametric indices. These patterns are consistent with inference-based noise and do not rely on parametric assumptions.

Biased PT estimations. KR state that "most binary choice datasets exhibit a correlation with the opposite sign to that reported in V24." This pattern can be traced to stochastic identification problems that arise when prospect theory (PT) parameters are estimated from binary choice data using random utility models. These problems take precisely the form described by Apesteguia and Ballester (2018), and are thus well well-known and well-understood.

One reason why I used certainty equivalents is that this approach mitigates these econometric issues. Writing the model as

$$ce = u^{-1}[w(p)u(x) + (1 - w(p))u(y)] + \epsilon,$$

where ϵ captures the residual, places the error term on the objective scale of monetary outcomes. This avoids the problem emphasized by Apesteguia and Ballester (2018), which arises from attaching independent error terms to utilities that are defined only up to arbitrary scale. When errors are attached to certainty equivalents, the relevant

scale is fixed by monetary units and is therefore no longer arbitrary. In binary choice, by contrast, this issue cannot be resolved as easily. Although Apesteguia and Ballester (2018) propose a solution for expected utility models, no robust solution is currently available for the estimation of PT functionals. As a consequence, PT estimates obtained from binary choice data using random utility specifications are systematically biased, particularly for highly risk-averse individuals. This bias is likely to drive the correlation patterns emphasized by KR.

To demonstrate that this is not merely a conjecture, I provide direct evidence using existing data. I analyze the first ten countries (in alphabetical order) from the dataset of L'Haridon and Vieider (2019), restricting attention to ten for computational reasons, as binary choice models are time-intensive to estimate. These data were collected using binary choices grouped into choice lists, with instructions enforcing single switching within each list. The same data can be analyzed in two ways: by attaching errors to certainty equivalents, or by estimating a random utility model on the underlying binary choices.

When errors are attached to certainty equivalents, the correlation between estimated likelihood sensitivity $\widehat{\gamma}$ and the noise parameter $\widehat{\sigma}$ is $\rho = -0.172$ with $p < 0.001$. Estimating a random utility model on the same underlying choices yields a correlation of $\rho = 0.196$ with $p < 0.001$. The same reversal occurs at the level of individual countries. Reversing the data generation process shows that similar patterns can arise in binary choice data when switching points are approximated in different ways. These results indicate that the positive correlations highlighted by KR are not driven by properties of the data themselves, but by well-known econometric issues associated with random utility estimation of PT parameters.

A further illustration comes from simulations based on the BIM. When BIM parameters are simulated, the response noise

$$\omega \triangleq \nu\sqrt{\gamma^2 + \alpha^2}$$

is negatively correlated with likelihood discriminability γ. This relationship underlies the prediction in Vieider (2024), since coding noise ν enters the definition of both quantities. The magnitude of the correlation depends on the signal-to-noise ratio and is therefore data-dependent in a straightforward sense. However, when a PT model is estimated on these same simulated data using a random utility specification, the estimated correlation between $\widehat{\gamma}$ and the noise standard deviation $\widehat{\sigma}$ again becomes positive. This reversal reflects mis-specification of the econometric model rather than any underlying feature of the data-generating process. In particular, high levels of risk aversion tend to be under-estimated, while likelihood sensitivity is over-estimated. The optimism parameter $\widehat{\delta}$ likewise under-estimates risk aversion relative to the weighted contribution of the prior mean, δ. These patterns underscore the difficulty of recovering meaningful PT parameters from binary choice data using random utility specifications.

Constructive directions. Further investigation of correlation structures in stochastic choice is clearly worthwhile. Progress on this question requires careful attention to both conceptual and econometric foundations. One promising direction is to focus on certainty-equivalent data, where identification issues are less severe. Extending such analysis to binary choice would require new solutions to longstanding identification prob-

lems in PT-based RUMs. While challenging, this line of research could yield valuable insights.

5 Conclusion

KR argue that the Bayesian inference model (BIM) of Vieider (2024) is not identified, and they derive several empirical claims from this premise. In this reply, I have shown that the claim of non-identifiability is incorrect. Once identifiability is established, the empirical conclusions drawn from its alleged absence no longer follow.

I have further examined the additional arguments advanced in the comment, including claims concerning differences between risk and mirrors, correlations between prospect-theory parameters, and utility under risk and ambiguity. While these topics are in principle worthy of further investigation, the analyses presented by KR do not provide new evidence that would substantively advance the debate. Several of the reported patterns can be traced to conceptual or econometric issues that are well understood in the literature.

More broadly, the discussion highlights the importance of distinguishing between descriptive random-utility representations and models of stochastic choice grounded in Bayesian inference. Confounding these frameworks leads to incorrect conclusions regarding identification, estimation, and empirical content. Careful attention to these distinctions is essential for progress in understanding stochastic choice behaviour.

References

Abdellaoui, Mohammed, Aurélien Baillon, Lætitia Placido, and Peter P. Wakker (2011) 'The Rich Domain of Uncertainty: Source Functions and Their Experimental Implementation.' *American Economic Review* 101, 695–723

Apesteguia, Jose, and Miguel A Ballester (2018) 'Monotone stochastic choice models: The case of risk and time preferences.' *Journal of Political Economy* 126(1), 74–106

Balcombe, Kelvin, and Iain M Fraser (2025) 'Flexible estimation of parametric prospect models using hierarchical bayesian methods.' *Experimental Economics*

Banki, Daniel, Uri Simonsohn, Robert Walatka, and George Wu (2025) 'Decisions under risk are decisions under complexity: Comment.' *Available at SSRN 5127515*

Barretto-García, Miguel, Gilles de Hollander, Marcus Grueschow, Rafael Polania, Michael Woodford, and Christian C Ruff (2023) 'Individual risk attitudes arise from noise in neurocognitive magnitude representations.' *Nature Human Behaviour* 7(9), 1551–1567

Borenstein, Michael, Larry V Hedges, Julian Higgins, and Hannah R Rothstein (2009) 'Introduction to meta-analysis'

Bouchouicha, Ranoua, and Ferdinand M. Vieider (2017) 'Accommodating stake effects under prospect theory.' *Journal of Risk and Uncertainty* 55(1), 1–28

Bouchouicha, Ranoua, Ryan Oprea, Ferdinand M. Vieider, and Jilong Wu (2024) 'Is prospect theory really a theory of choice?' Technical Report, Ghent University Discussion Papers

Bouchouicha, Ranoua, Yuchi Li, and Ferdinand M. Vieider (2025) 'Loss-sensitivity versus loss-aversion.' Technical Report, Ghent University Discussion Papers

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer (2024) 'Meta-analysis of empirical estimates of loss aversion.' *Journal of Economic Literature* 62(3), 485–616

de Hollander, Gilles, Marcus Grueschow, Franciszek Hennel, and Christian C. Ruff (2024) 'Rapid changes in risk preferences originate from bayesian inference on parietal magnitude representations.' *bioRxiv*

Fehr-Duda, Helga, Adrian Bruhin, Thomas F. Epper, and Renate Schubert (2010) 'Rationality on the Rise: Why Relative Risk Aversion Increases with Stake Size.' *Journal of Risk and Uncertainty* 40(2), 147–180

Gelman, Andrew, and Hal Stern (2006) 'The difference between significant and not significant is not itself statistically significant.' *The American Statistician* 60(4), 328–331

Herold, Florian, and Nick Netzer (2023) 'Second-best probability weighting.' *Games and Economic Behavior* 138, 112–125

Hogarth, Robin M., and Hillel J. Einhorn (1990) 'Venture Theory: A Model of Decision Weights.' *Management Science* 36(7), 780–803

Imai, Taisuke, Salvatore Nunnari, Jiling Wu, and Ferdinand M. Vieider (2025) 'Meta-analysis of prospect theory parameters.' *Working Paper*

Khaw, Mel Win, Ziang Li, and Michael Woodford (2021) 'Cognitive imprecision and small-stakes risk aversion.' *The Review of Economic Studies* 88(4), 1979–2013

_ (2023) 'Cognitive imprecision and stake-dependent risk attitudes.' Technical Report

L'Haridon, Olivier, and Ferdinand M. Vieider (2019) 'All over the map: A worldwide comparison of risk preferences.' *Quantitative Economics* 10, 185–215

Ma, Wei Ji, Konrad Paul Kording, and Daniel Goldreich (2023) *Bayesian Models of Perception and Action: An Introduction* (MIT press)

McElreath, Richard (2016) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Academic Press)

Natenzon, Paulo (2019) 'Random choice and learning.' *Journal of Political Economy* 127(1), 419–457

Netzer, Nick, Arthur Robson, Jakub Steiner, and Pavel Kocourek (2024) 'Risk perception: measurement and aggregation.' *Journal of the European Economic Association* p. jvae053

Oprea, Ryan (2024) 'Decisions under risk are decisions under complexity.' *American Economic Review*

Oprea, Ryan, and Ferdinand M. Vieider (2024) 'Minding the gap: On the origins of the description-experience gap.' *Working Paper*

Oprea, Ryan, and Ferdinand M Vieider (2025) 'The modern psychophysics of risk and time.' *Working Paper*

Steiner, Jakub, and Colin Stewart (2016) 'Perceiving prospects properly.' *American Economic Review* 106(7), 1601–31

Van Bergen, Ruben S, Wei Ji Ma, Michael S Pratte, and Janneke FM Jehee (2015) 'Sensory uncertainty decoded from visual cortex predicts behavior.' *Nature neuroscience* 18(12), 1728–1730

Vieider, Ferdinand M. (2024) 'Decisions under uncertainty as Bayesian inference on choice options.' *Management Science* 70, 9014–9030

_ (2026) 'The Structure of Stochastic Choice.' Technical Report, Ghent University Discussion Papers