Referee Report:
Comment on "Decisions under Uncertainty as Bayesian Inference on Choice Options"

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.

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 the BIM 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 seems baffling in light of the fact that I explicitly introduce parameter restrictions with the aim of guaranteeing identifiability. The parameter definitions, shortly repeated in the summary above, clearly show that the assumption of common coding noise creates a common scale for the two discriminability parameters α and γ. These parameters are fully defined by the signal to noise ratios σ_o/ν and σ_p/ν, and thus measured on the scale of the common coding noise ν. Using simulations it is straightforward to show the base truth — parameter recoverability is, in fact, excellent.

Given that the parameters are identified, the claims on mirrors versus risk do not follow. KR however make additional allegations in this instance. One claim is that their estimations based on a restricted version of Vieider's BIM leads to "different conclusions" (p. 2). The conclusions they present, however, merely repeat statements already contained in Vieider (2024). 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). It is however easy to show that changing the priors from being mildly informative to improper (i.e. governed by the likelihood alone), does not produce *any changes* in estimated parameters. In light of this, the claim made by KR is truly surprising.

The claims on utility differences under risk and ambiguity largely miss the point of the exercise. Although a closer investigation of this issue could have been interesting, KR bring no new evidence to the table. They take data obtained from a bisection procedure — which cannot be described by the BIM — and add them to the original data. They then search for additional tests based on which the results seem to fall short of traditional thresholds of significance. The thereby fall pray to common statistical fallacies. Most importantly, the key point of this experiment is not addressed: the specific instance of violations of probability-outcome separability is just one in a list of such violations in the literature. Vieider (2024) explicitly presents the whole list as being predicted by the BIM—a fact that is neither acknowledged, nor addressed in the comment.

Finally, KR revisit correlations between likelihood-discriminability $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$ and the standard deviations of response errors, $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$. Their point again builds on their allegation of lack of identifiability, which in this instance takes the form of the BIM allegedly not making any quantitative predictions on the correlation. It is easy to show that this claim, too, is false. Conceptually, it is hard to reconcile with the very precise predictions emerging from model-based simulations. The so-called "meta-analysis" — consisting in the aggregation of a handful of cherry-picked studies — further runs into well-known issues pertaining to the econometric recoverability of prospect theory parameters in random utility models (which the BIM solves).

In sum, the key claims of lack of identifiability, results being driven by priors, and the model not making any quantitative predictions on correlations between noise and discriminability, *can easily be shown to be false*. Below, I will show at some length how a careful examination of the issues establishes this beyond reasonable doubt. Some of the empirical exercises — particularly on utility differences, and on correlations — would in principle be worth conducting. At present, however, the comment does not bring any new evidence to the table.

1 Identifiability and number of parameters

The key claim KR make concerns identifiability of the BIM parameters. A closely linked secondary claim is that the BIM has too many parameters. Here, I first show that KR's claim of lack of identifiability is factually incorrect. I then argue that the counting of parameters and comparisons to PT are not meaningful in this context, because *the parameters 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 seem mystifying in light of that fact that 1) in Vieider (2024), I explicitly justify modelling coding noise as common to log-odds and log cost-benefits in order to guarantee identifiability ("the coding noise is modeled as being common to the two dimensions because different coding noises would not be separately identifiable from direct tradeoffs", p. 9018); and 2) the definitions $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$, and $\alpha \triangleq \frac{\sigma_o^2}{\sigma_o^2+\nu^2}$ clearly put the two parameters on the common scale of coding noise ν. The common coding noise assumption is isomorphic with the correlation structure used by Natenzon (2019) to guarantee identifiability of the multinomial Probit.

In case of any residual doubt, it is simple to establish the base truth by testing the econometric recoverability of choice patterns simulated from the model. To do this, I 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 coding noise ν using simulations. I simulate 100 subjects using the following simulated parameters:

$\nu \sim \left|\mathcal{N}(0.6, 0.3)\right|$

$\sigma_p \sim \left|\mathcal{N}(0.7, 0.35)\right|$

$\sigma_o \sim \left|\mathcal{N}(0.9, 0.4)\right|$

$\xi \sim \left|\mathcal{N}(0.7, 0.3))\right|$.

The draws are taken independently for each parameter. The distributions are such that i) there is a reasonably high signal-to-noise ratio on average (i.e. $\sigma_p/\nu > 1$ and $\sigma_o/\nu > 1$ on average); but that ii) there is high variation in the supposedly non-identified parameters, which makes the recovery exercise meaningful and maximizes power to detect any (absence of) correlations. I subsequently estimate the BIM putting non-informative hyperpriors on the population level parameters using a normal distribution with mean 0 and SD 10 (I will show below that, in any case, the estimation results are not sensitive to the choice of priors within reasonable limits). Since these hyperpriors are specified on a log scale, they imply priors that attribute 95% of the probability mass to the interval between 0 and $exp(1.96 * 10) \approx 3.25 * 10^8$ — a range that for all practical purposes may as well be infinite. The econometric code is otherwise the same as used by Vieider (2024). I use rich stimuli incorporating 300 binary choices by simulated subject, with orthogonal variation in the log-odds and log cost-benefits.

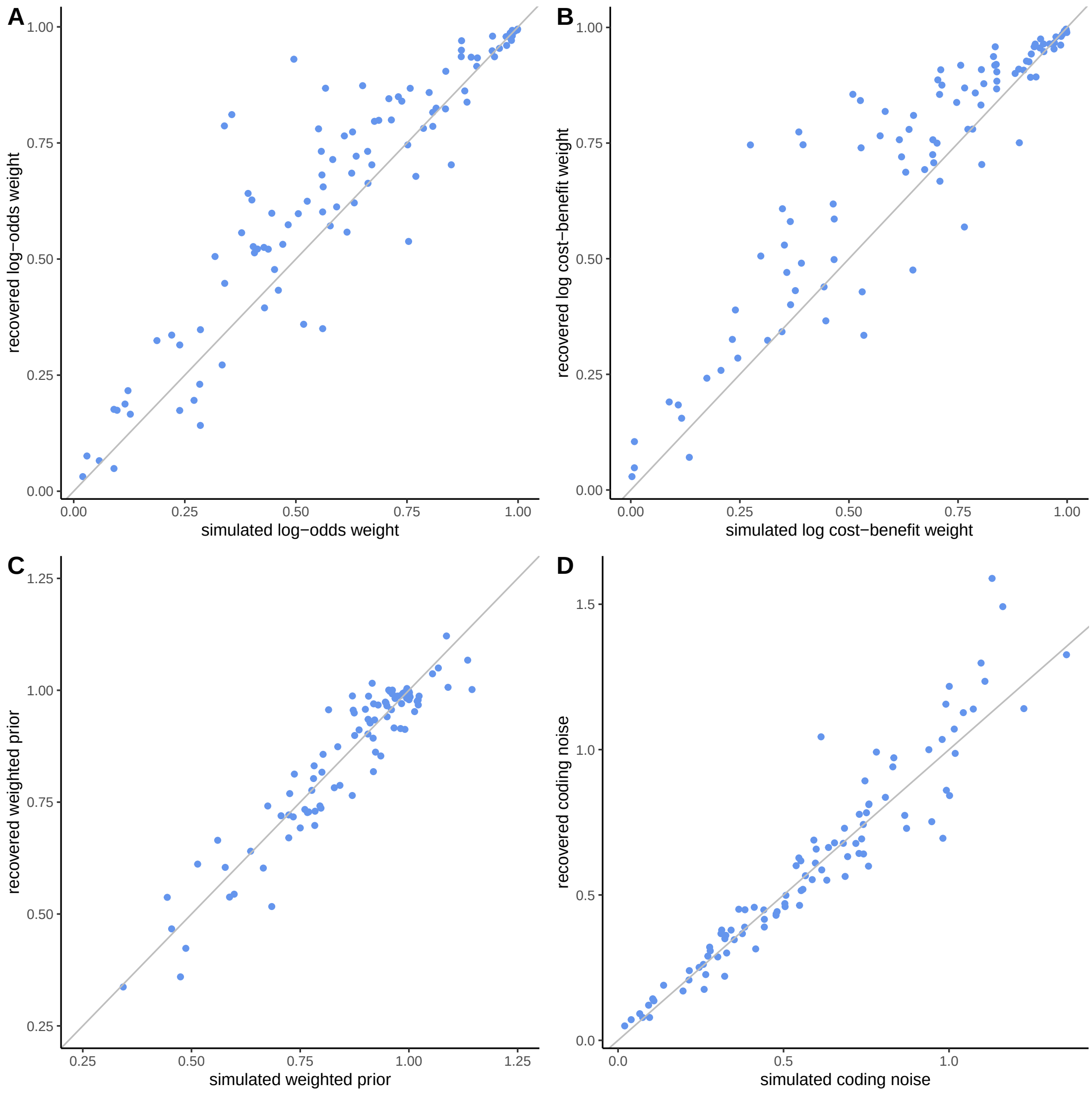

Figure 1: Scatter plots of simulated vs recovered parameters

The figures shows correlations between simulated and recovered parameters: likelihood-discriminability γ in panel A; outcome-discriminability α in panel B; and the weighted prior mean δ in panel C. Panel D shows the recoverability of coding noise ν. Continuous gray lines indicate the point of parameter equality, i.e. they are 45 degree lines.

Contrary to what KR claim ("it is impossible to separate α from γ", p. 3), the recoverability of parameters is excellent. Figure 1visually displays the correlations between the simulated parameters and the parameters recovered from the econometric estimations (all Spearman correlations > 0.9). They can be seen not only to be closely correlated, but also to have the same magnitudes These are correlations towards the upper limit of what is reasonably possible based on noisy data. It is thus hard to see where KR's claim of lack of identifiability may come from.

Number of parameters. A secondary complaint by KR concerns the number of parameters in the BIM. Here, the issue seems to stem from an implicit assumption that generative parameters are equal in standing to descriptive parameters — they are not. I briefly discuss this issue here as a basis for what comes below.

KR claim that there are 5 parameters in the model. Vieider (2024) uses only 4 in the estimations. Whatever the exact number, this discussion misses the central point of the model: as repeated emphasized in Vieider (2024), PT parameters are *descriptive* in nature, while the BIM parameters are *generative*. In a dynamic environment, separate priors are key variables for predicting choice, as Bouchouicha et al. (2025) illustrate based on a 6-parameter version of the model of Khaw et al. (2021).

Let me start by discussing the number of parameters in prospect theory (PT). An estimation of PT on pure gain lotteries can be parameterized in different ways. One could use a 3-parameter specification (1 for utility curvature, 1 for probability distortions, 1 for noise); a 4-parameter specification (adding an additional parameter for the elevation of the probability weighting function); a 5-parameter specification (e.g. adding an additional parameter for an expo-power utility function); or indeed a 6-parameter specification (e.g. adding an additional 'tremble' term to the error).[1] The choice of the number of parameters is typically determined after observing the data to obtain the best possible fit. The parameters are independent of each-other, and are only mildly constrained (e.g, some of the parameters need to be strictly positive for the functions to be defined).

KR implicitly assume the parameters of the BIM to be of the same type. i.e. to be meant for data fitting. This, however, does not correspond to the spirit of the model. Given the *generative* nature of the model, the parameters should be interpreted as *causal*: they receive their meaning from the mental information processing they are supposed to guide. That is, they pre-exist choice, and causally determine choices through the mental inference process about the choice primitives presented by the experimenter, which are modelled as latent and thus not directly accessible to the decision-maker. A fully-fledged model *could* thus include the 5 parameters mentioned by KR. It could even include six parameters, e.g. separating coding noise to be specific for outcomes (call the standard deviation ν_o) or to probabilities or events (call the standard deviation ν_e). This does not mean that all 6 parameters need to be recoverable from binary choices, as KR assume (though the 4 parameters in the journal version are explicitly chosen to be econometrically recoverable, as proven 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 behaviour (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).

Several recent papers further illustrate the implications of this generative interpretation. In Oprea and Vieider (2024), standard probability dependence in choice is established

[1] To illustrate that these examples are practically relevant, a recent estimation by Balcombe and Fraser (2025) used a 17-parameter version to fit choices over gains, losses, and mixed gambles, with some parameters explicitly added to obtain parameter values that agree with 'typical values' one should observe accoding to PT.

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 such as PT, 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 the signal-to-noise ratio 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 behavioural 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 identifiability. KR's claim on the lack of identifiability of BIM parameters is clearly untenable. I now turn to their 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 priors adopted in the Bayesian estimations. 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 behaviourally meaningful — to wit, I explicitly discuss it in my paper (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 behaviour within choice lists, does not apply to binary choice data. In particular, my 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. In light of the identification of the full model proven above, the claim by KR that parameter estimates published in the paper are only due to restrictive priors seem truly surprising. Luckily, we do not need to reply on intuition here — the claim is exceedingly easy to test.

The priors in Vieider are defined on the population-level parameters and are thus truly *hyperpriors*. They are chosen to be mildly informative to help convergence of the estimation algorithm — something that is standard practice in Bayesian econometrics (e.g. McElreath, 2016). Given the amount of data, any major effect on parameters seems unlikely. The statement of 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) is easily put to the test. To this end, I 1) ran the version of the BIM underlying the published results in Vieider (2024); and 2) ran the same model with 'improper hyperpriors', i.e. setting the priors equal to the entire support of the parameters, equivalent to a frequentist estimate.

Figure 2 shows scatter plots of the parameters recovered from the restricted version (on the x-axis), with parameters from the frequentist equivalents (on the y-axis). The correlations are almost perfect, and show no trace of "estimates [that] take arbitrary values driven by priors". The foundation for this claim is thus once more baffling.

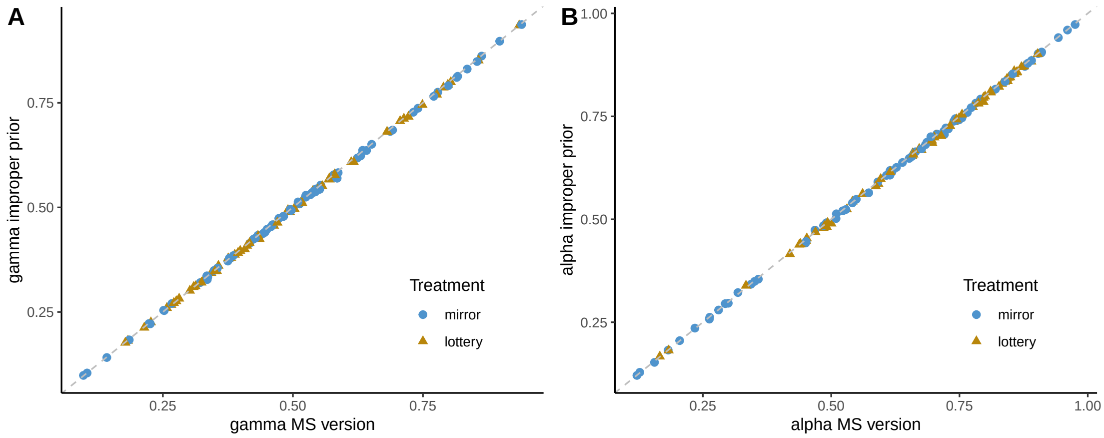

Figure 2: Scatter plots of parameters with informative versus improper priors
Scatter plot of parameters estimated using the code used by Vieider (2024), and parameters recovered from an equivalent model with improper hyperparameters ranging over the entire parameter support. The latter are equivalent to a frequentistic estimate of the model. Panel A show a scatter plot for likelihood-discriminability γ, and panel B for outcome-discriminability α. The correlations are almost perfect in both cases.

Conclusion on lotteries vs mirrors. It is hard for me to reconstruct where the claims in the comment may come from. KR apply what they call a "new version" of the BIM, even though this version of the model is shown in the paper. They claim different conclusions because they document some differences between risk and mirrors in their parametric analysis, even though these exact differences are already discussed in the paper. Their claim on parameter estimates obtaining purely based on restrictive priors seems even more surprising, given that it is not backed up by any evidence whatsoever.

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.[2]

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 favour 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 white noise and likelihood-sensitivity

The part on the correlation of likelihood-sensitivity and noise in PT models is potentially the most interesting aspect in the comment. The present comment, however, 1) cherry-picks a handful of ill-defined studies, many of which do not fall under the remit of the BIM; 2) relies on the same conceptual assumptions that underlie earlier claims regarding identifiability and priors; and 3) does not take into account well-known difficulties in recovering prospect-theory (PT) parameters from binary choice data using random utility models (RUMs). I discuss these 3 points in turn.

Meta-analysis and study selection and definition. Several of the studies in the aspirationally named "meta-analysis" use bisection methods — to the contrary of what KR state in their note — and can thus not be meaningfully described by the BIM. Such bisection requires complex explanations, is not incentive-compatible, and will trigger different cognitive processes than either binary choice or CEs (see meta-analysis by Imai et al., 2025, for evidence that bisection produces systematically different estimates of

[2]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).

PT parameters from choice and CEs). It is also unclear how they define a "study", their prime unit of analysis: they (i) focus only on gains, and drop losses, which are included in Vieider (2024); and (ii) present the 3 different experiments in Bruhin et al. (2010) as 3 separate "studies", but the 30 experiments in L'Haridon and Vieider (2019) as only 1 "study". In their graphs in appendix D, they further drop the estimates from Bruhin et al. (2010) all together (although the title suggests the opposite). The method underlying the meta-analysis is not described (not even in the appendix). Is this a fixed-effects or a random-effect analysis? How do they deal with findings that lack statistical independence? Are there controls or adjustments for publication bias? None of these crucial issues are discussed (see e.g. Brown et al., 2024, on ways to tackle some of these issues in a principled way).

A true meta-analysis on correlations between noise and likelihood-sensitivity estimated based on certainty equivalents could indeed be useful here. It would also be very easy to carry out. Bouchouicha et al. (2024) provide a full list of all papers ever estimating PT functionals from CEs, as well as based on pure binary choice. A hallmark of meta-analysis is indeed that *it needs to include all available data* (Borenstein et al., 2009). First, however, understanding how to do this requires putting some concepts in place. The discussion here in sequences follows Vieider (2026), where I examine the stochastic choice predictions of Bayesian inferences models from both a theoretical and empirical perspective.

Conceptual problems. KR make strong claims about the BIM's predictions about correlations between likelihood-discriminability $\gamma \triangleq \frac{\sigma_e^2}{\sigma_e^2+\nu^2}$ and the response error $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$. They claim, for instance, that "it is impossible to predict the magnitude of this correlation" (p. 3). This claim, in turn, is closely linked to identifiability problems in their econometric approach on binary choice data (**NOTE**: in the main text, they state that they use observed CEs and compare them to theoretical CEs also for binary choice, but that it impossible; the referenced technical appendix shows that they actually use a random utility model, thus defining the error for certainty equivalents and in binary choice on different scales). I first discuss conceptual issues at some length, before turning to the econometric issues.

KR state that "when [the] BIM predicts a correlation between PT parameters, it is impossible to predict the magnitude of this correlation" (p. 3), and that "Understanding why estimations based on CE [sic!] measured with choice lists generate stronger $\widehat{\gamma}.\sigma$ correlations is beyond the scope of this note". Both statements are surprising. The magnitude of the correlation within the BIM is exceedingly simple to predict: For a given cost-benefit signal-to-noise ratio σ_o/ν, it is simply a function of the log-odds signal-to-noise ratio ν/σ_p . As one can see from the quadratic form of the response variance (see **?** for details), for given prior variances σ_p and σ_o the response error is a non-monotonic function of coding noise ν. This is indeed a key feature of the BIM, which means that the predicted mappings from choices to average choice frequencies is monotonic. The model shares this key feature with the model of Khaw et al. (2021). This is a textbook case of the properties of Bayesian inference: Ma et al. (2023), cited by Vieider, provide the intuition and proof in their chapters 3 and 4.

Figure 3 shows the evolution of the response error, $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$ as a function of coding noise ν for when $\sigma_p > \sigma_o$ (in the example, I use the means used in the simulations

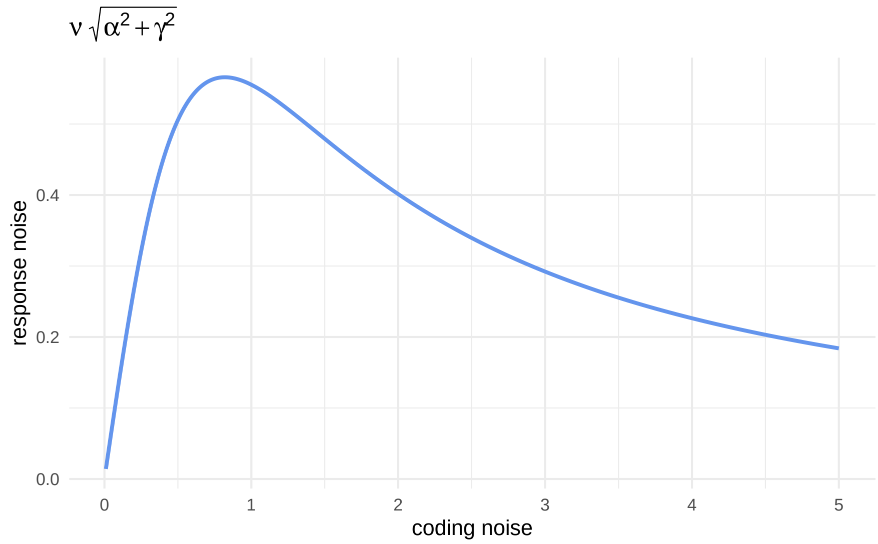

Figure 3: Response noise as a function of coding noise

The figure illustrates the non-monotonic evolution of response noise, $\omega \triangleq \nu\sqrt{\alpha^2+\gamma^2}$, as a function of coding noise ν in the BIM. This feature ensures that the model — to the contrary of the random utility model — is stochastically monotonic in the level of risk aversion manifested in choices.

above). The intuition for the non-monotonicity of response noise as a function of coding noise is simple: when the signal-to-noise ratio becomes unfavourable, a decision-maker will increasingly rely on the prior, and behaviour will thus become *less* variable. Ma et al. (2023) provide an entry-level explanation and proof in chapters 3 and 4 based on a uni-dimensional case. Response noise rises rapidly at first, but then gradually declines again after reaching its peak. As coding noise increases, choice behaviour will thus slowly converge towards the mean of the prior.

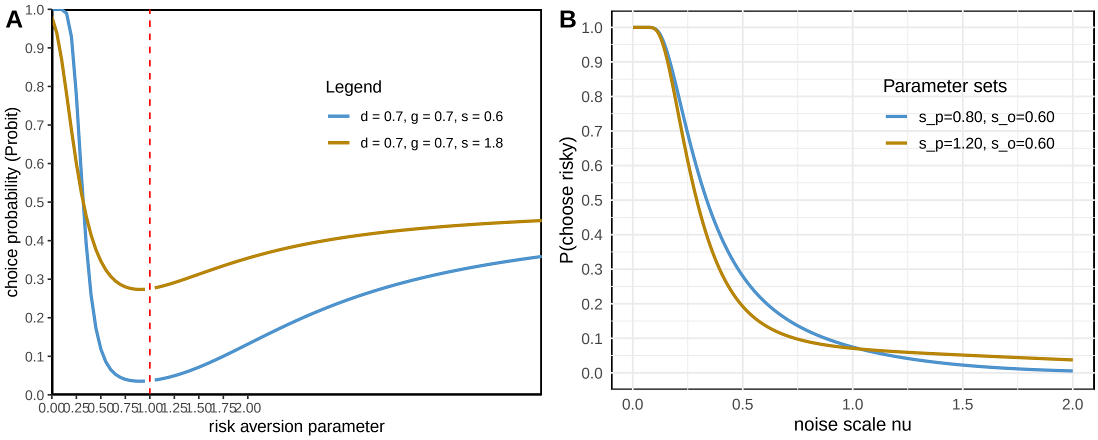

Figure 4: Response noise as a function of coding noise

The figure illustrates stochastic parameter recoverability in the RUM and the BIM. The patterns are illustrated using a lottery providing a 10% chance at \$60, or else \$1, compared to a sure outcome of \$ 5 (example from Apesteguia and Ballester, 2018). As utility curvature increases in the CRRA function, choice proportions of the risky option decrease at first, but then slowly revert towards 50-50. Parameters in the legend are as follows adopting the notation from Vieider: $d=\widehat{\delta}$, $g=\widehat{\gamma}$, and $s=\widehat{\sigma}$. Apparent discountinuities at $\widehat{\rho}=1$ in the CRRA function are an artifact of the figure, since utility for $\widehat{\rho}=1$ should be $ln(x)$ for the function to be defined. Panel B shows comparable figures for the BIM as coding noise ν increases. Other than in the PT RUM, choice patterns are monotonic. The parameters are as follows: $s\_o=\sigma_o$, $s\_p=\sigma_p$.

All of this is exceedingly simple to explore by simulation. Figure 4, panel A, shows a simulation of stochastic choice proportions of the lottery predicted a random utility Probit model based on underlying PT parameters.[3] As utility curvature or risk aversion increases, predicted choice proportions of the lottery first decline, but later rebound. This creates the well-known issues in the recoverability of PT parameters using random utility models (Apesteguia and Ballester, 2018). Panel B simulates choice proportions in the BIM as a function of coding noise. Choice proportions of the lottery now monotonically increase in coding noise ν, showing how the BIM overcomes issues in random utility models estimating PT parameters. This topic is treated in depth by Vieider (2026).

Possible differences between certainty equivalents and binary choices can emerge because binary choice has a higher signal-to-noise ratio. This is indeed shown directly by Bouchouicha et al. (2024), and predicted by their generalization of the BIM to choice list data. In particular, using the universe of PT estimates based on binary choice or certainty equivalents, Bouchouicha et al. (2024) show that binary choice yields consistently higher levels of likelihood-sensitivity $\widehat{\gamma}$. Supplementing this evidence with several experiments, they further show that this goes hand-in-hand with decreased noise replative to certainty equivalents. Although switching towards the middle of the list could play a role, as suggested by KR, Bouchouicha et al. (2024) test a distinctive prediction of their cognitive model that predicts list-like patterns to appear in binary choice when options are presented sequentially, and find this prediction to be supported.

Biased PT estimations. KR state that "most binary choice datasets exhibit a correlation with the opposite sign to that reported in V24". This problem can be traced to a stochastic identification issues of PT parameters estimated based on random utility models, which take exactly the form of the well-known Apesteguia and Ballester (2018) critique. Tis creates non-trivial problems for the estimation of PT parameters, given that the mapping between choice proportions and parameters governing risk-taking in PT is not unique.

The reason for which I used certainty equivalents in my paper is that this allowsme to overcome these econometric issues. Writing the model as $ce = u^{-1}\left[w(p)u(x) + (1 - w(p)u(y))\right] + \epsilon$ (where ϵ captures the residual) puts the error on the objective scale of monetary outcomes. This overcomes the problem raised by Apesteguia and Ballester (2018), which arises from independent arbitrary scales in the numerator and denominator when errors are attached to utilities. Now, the scale in the numerator is the objective scale of monetary amounts, and thus no longer arbitrary. In binary choice, this issue cannot easily be overcome. Although Apesteguia and Ballester (2018) propose a solution to this issue in an expected utility model, I am not aware of any robust solutions when it comes to the estimation of PT functionals. The PT estimations presented by KR are thus systematically biased at least for part of the subjects, which is likely to drive the patterns they show.

To demonstrate that this is not merely a conjecture, I take the first 10 countries (in alphabetical order) from the data of L'Haridon and Vieider (2019) to illustrate this point (restricting attention to ten for computational reasons). The data were collected using

[3] In a PT setting such as applied here, this issues does not only occur as a function of increasing utility curvature. It could, for instance, also occur in the comparison of two lotteries if the elevation of the probability weighting function becomes more and more depressed. Practically, the two elements are notoriously difficult to separate in estimations.

binary choices grouped into choice lists, with the instructions enforcing single switching in a given list (I do not have access to the Bruhin et al. (2010), data in both formats). The data can be analyzed in two different ways — using errors attached to the certainty equivalents, or estimating a random utility model on the underlying binary choices. If I examine the correlations between likelihood-sensitivity $\widehat{\gamma}$ and the noise term $\widehat{\sigma}$ for the specification based on certainty equivalents, the correlation is $\rho = -0.172, p < 0.001$. If, however, I examine the exact same correlation based on the random utility specification based on binary choices, the correlation is $\rho = 0.196, p < 0.001$. The same reversal also occurs in the single country data. This shows that the positive correlations shown by KR are *not due to data*. Instead, they are due to well-known problems in the econometric estimation of random utility models.

There is yet another way of showing this, by using the simulations I already used above for identification. When I simulate the BIM parameters, the response noise $\omega \triangleq \nu\sqrt{\gamma^2 + \alpha^2}$ is negatively correlated with the likelihood-discriminability parameter γ. This is indeed the basis for the prediction in Vieider (2024): after all, coding noise ν enters the definition of both. The quantitative strength of the correlation will depend on the signal-to-noise ratio, and in this sense it is indeed 'data-dependent' — a trivial observation. In Vieider (2026), I indeed manipulate the signal-to-noise ratio in the outcome dimension, and find support for the inverse-U shaped relationship between outcome-discriminability and decision noise predicted by the BIM (and more generally by models of Bayesian inference).

Estimating a PT model using a random utility model on these same simulated data now again produces a *positive* correlation between likelihood-sensitivity $\widehat{\gamma}$ and the noise standard deviations $\widehat{\sigma}$. This shows how these patterns emerge from the mis-specification of the econometric model, rather than anything real pertaining to the data. The most severe levels of risk aversion are under-estimated, and on the flipside, likelihood-sensitivity is over-estimated. The optimism parameter $\widehat{\delta}$ systematically under-estimates risk aversion compared to the weighted contribution from the prior mean, δ. This yet again shows the difficulty in recovering meaningful parameters using random utility specifications to estimate PT parameters.

Constructive advice. Further studying the correlation structure underlying PT parameters is certainly a worthwhile endeavour. To investigate the issue in binary choice, however, would first require developing solutions to RUM identifiability issues applicable to PT parameters. While his would inevitably be a difficult task, it also comes with high rewards, and I can only encourage it.

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

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010) 'Risk and Rationality: Uncovering Heterogeneity in Probability Distortion.' *Econometrica* 78(4), 1375–1412

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