Evaluating d': Global Motion Detection requires Random Scale Representation and Allows for Threshold Representation

Ruby de Lanerolle (ruby.delanerolle.22@ucl.ac.uk)[1] **& Henrik Singmann (singmann@gmail.com)**[1]
[1]Department of Experimental Psychology, University College London

Abstract

Global motion detection performance is often described via d' (i.e., equal-variance Gaussian Signal Detection Theory), despite evidence that global motion detection requires a threshold representation which is incompatible with d'. Both d' and threshold representation share the core assumption of random-scale representation (RSR), which posits a latent evidence scale without committing to specific distributional forms. We apply critical tests to assess whether global motion detection requires a RSR and whether a threshold representation is admissible. Participants completed an m-alternative forced-choice random dot motion task with $m = \{2 \dots 8\}$. Accuracy across m satisfied RSR constraints, providing strong evidence that global motion detection relies on RSR. We derived a hazard function from the accuracy and tested the monotonicity constraint implied by threshold representation. The hazard function satisfied monotonicity, consistent with a threshold account. Reconstructed ROC curves were piece-wise linear, inconsistent with equal-variance Gaussian SDT, while generally consistent with threshold models.

Keywords: global motion detection; random-scale representation; threshold model; signal detection theory

Introduction

Imagine you are looking at a tree and there is a light breeze rustling through its leaves. The leaves flutter about, not all moving at the same time nor in the same direction. However, at any given time, you see that some of the leaves are moving in a single direction, allowing you to infer which way the breeze is blowing. This ability to detect the direction of motion from noisy visual information is known as global motion detection.

The typical lab analogue for global motion detection is the random dot motion paradigm illustrated in Figure 1A (e.g., M. L. Green & Pratte, 2022, 2026; Shadlen & Newsome, 1996; Williams & Sekuler, 1984). In random dot motion experiments, participants are shown apertures containing many moving dots. Some dots move in a single direction, while others move randomly. The dots moving in the same direction seem to form a transparent screen of motion, appearing as global motion across the aperture (Schütz et al., 2010). Here, we are primarily interested in versions of the task in which participants need to make a binary decision: Is there coherent motion in the aperture or only random noise?

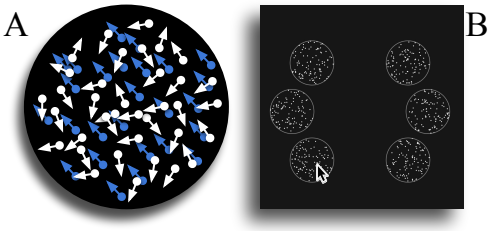

Figure 1: A: Illustration of a random dot motion aperture containing global motion. Some dots (highlighted in blue) move in a single direction, while the rest move randomly. B: Example of 6-AFC global motion detection stimuli. One of the $m = 6$ apertures contains some coherent motion, the rest contain only noise. Participants click to select the signal.

Signal Detection Theory and Random-Scale Representation

One prominent formal approach for understanding decision-making in global motion detection tasks is Signal Detection Theory (SDT; D. M. Green & Swets, 1966; Kellen & Klauer, 2018). According to SDT, global motion decisions rely on a latent perceptual strength formed by pooling noisy evidence for coherent motion. Internal noise may further perturb this evidence. SDT further assumes that apertures containing coherent motion – signal stimuli – and those containing only random motion – noise stimuli – are described by different distributions on a latent strength scale. Depending on the task, decision rules may vary. For example, if observers only judge whether coherent motion is present or absent – a yes-no task – the latent strength is compared against a response criterion. If the observer has to select the stimulus containing coherent motion among noise alternatives – an m-alternative forced-choice task (m-AFC) – the aperture with the highest latent strength is selected.

In the global motion detection literature, as in many other domains, SDT is most commonly applied in its equal-variance Gaussian parametrisation, typically via the use of the discriminability index d' (e.g., Barlow & Tripathy, 1997; Britten et al., 1992; Desender et al., 2021, 2022; Lappin & Bell, 1976; Liang et al., 2016; Newsome et al., 1989). Yet the Gaussian assumption itself has rarely been tested, leaving open whether the widespread use of d' in this domain is justified.

One direct test of the Gaussian assumption comes from Liang et al. (2016). They compared d' estimates across two

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

direction-detection tasks (2-AFC and same-different) using the same apertures. All apertures had 100% motion coherence, (i.e., all dots moving in the same direction) and so authors assumed all noise was internal. By the Generalised Area Theorem (Iverson & Bamber, 1997), d' should be self-consistent across tasks, and so if the evidence distributions were Gaussian, d' would be identical across the two tasks. Liang et al. found, rather, that d' differed across tasks, suggesting that at least once internal noise perturbs evidence distributions, they are no longer Gaussian.

Another result questioning the suitability of the Gaussian SDT model for global motion detection comes from a study by M. L. Green and Pratte (2022). In a random dot motion task with one aperture presented per trial, participants had to report the direction of coherent motion. Coherence (i.e., the percentage of dots moving in the same direction in the signal aperture) differed between trials. If global motion detection were a continuous process, congruent with continuous SDT-like models, then increasing coherence would be associated with a gradual increase in task performance, as pooling evidence from more dots should increase precision in direction-perception. Instead, they found data consistent with a mixture of highly accurate answers and pure guesses. Additionally, while accuracy increased with coherence, the precision with which accurate trials were answered was constant across coherence levels. Together these results support that global motion detection is governed by a threshold all-or-nothing process rather than the continuous evidence accumulation that a continuous (e.g., Gaussian) SDT model would require.

While SDT is often thought of and applied only in its equal-variance Gaussian parametrisation, it can in fact accommodate many different distributional assumptions (D. M. Green, 2020; Kellen et al., 2021). To understand the representational core of SDT, we can consider the equal variance Gaussian model and then strip away the parametrisation. For an m-AFC global motion detection task, each stimulus elicits a certain latent evidence strength. The strength value elicited by the aperture containing coherent motion is drawn from the Gaussian signal distribution, while the strength values of the $m - 1$ apertures containing random motion only are drawn from the Gaussian noise distribution. However, we can strip away the assumptions that these are equal-variance Gaussian distributions, and we would be left with the assumption that there are *some* distributions for signal and noise on the latent evidence scale. These distributions need not be independent, and they need not be the same. This minimal set of assumptions – that stimuli elicit evidence strengths which are sampled from distributions on a shared latent scale, and that decisions are made by comparing sampled strengths – is referred to as *random-scale representation* (RSR; Falmagne, 1978). In fact, as shown by Kellen et al. (2021), even threshold models imply a RSR, despite often being conceived of as fundamentally different from SDT models. In other words, threshold models are just SDT models with unusual evidence distributions.

Critical Tests of Random-Scale and Threshold Representations

Formally, RSR entails a set of order constraints on the probabilities of choosing the signal option when it is presented together with $m - 1$ noise alternatives in the m-AFC task, called the Block-Marschak Inequalities (Block & Marschak, 1974). These inequalities include at least chance performance, a monotone decrease in accuracy as m increases, the rate of decrease in accuracy decreasing with m, and a series of higher-order constraints that together tightly constrain the curvature of the accuracy sequence across m. Importantly, Falmagne (1978) proved that RSR is satisfied if and only if the Block-Marschak Inequalities are satisfied. Thus, satisfying the Block-Marschak inequalities is formally equivalent to the existence of a RSR.

The fact that RSR entails a testable set of constraints allows testing whether the assumptions also shared by SDT models are empirically viable. Instead of having to test particular instantiations of SDT, we can test a whole family of models without having to make specific distributional commitments. Applied to our domain of interest, the logic of the critical test is as follows: If the test for RSR passes, we know that there is some SDT-like model that can be used to describe global motion perception. If the test fails, any model sharing the same assumption – including equal-variance Gaussian SDT, d', and even threshold models – is inadmissible for describing global motion perception.

Recently, Kellen et al. (2021) tested the constraints implied by RSR in recognition memory. They first asked participants to learn a set of words. In the subsequent test phase participants were presented with m-AFC trials with m going from 2 to 8 in steps of 1. Results of two experiments showed that participants' responses were in line with constraints imposed by RSR providing strong evidence that recognition judgements in memory are described by an SDT-like model with unknown evidence distribution shape.

Despite the breadth of models included in the RSR family, the Block-Marschak Inequalities are a very strict set of constraints. Kellen et al. (2021) show that for an m-AFC task as used in the present study, when performance is above chance and steadily decreases with m, RSR is only satisfied in $1/413{,}121{,}934{,}659$ of the outcome space. Since this shows that RSR is by no means trivially satisfied, whether models relying on this assumption are admissible in global motion detection must be empirically tested.

The first aim of our paper is to test RSR in global motion detection. To the best of our knowledge this is the first paper to test whether all models which rely on RSR are feasible for global motion detection. We will use the same general approach as Kellen et al. and test whether judgements in an m-AFC motion detection task with $m = 2$ to $m = 8$ are consistent with the constraints implied by RSR. Furthermore, we follow the procedure of He et al. (2026) and include the test of likelihood-ratio monotonicity in our test of RSR. Likelihood-ratio monotonicity describes the constraint that

the likelihood of a signal being detected can only increase as evidence strength increases.

Another contribution of Kellen et al. (2021) and following work (He et al., 2026; Meyer-Grant et al., in press) is the dismissal of threshold representations in recognition memory in favour of continuous-strength models. In global motion detection, on the other hand, M. L. Green and Pratte (2022) provided the aforementioned evidence for a threshold representation over continuous-strength models. Thus, the second goal of the present paper is to test for threshold representation in global motion detection. To this end, we will perform the critical test for threshold representation recently introduced by Chechile and Dunn (2021; for an empirical example see McCormick & Semmler, 2023).

Chechile and Dunn's (2021) test is based on data from a multi-item ranking task. Just as in an m-AFC task, each trial of a multi-item ranking task presents participants with m items, one signal item and $m - 1$ noise items. However, instead of only selecting one item as in an m-AFC task, participants are asked to rank the items according to their perceived evidence strength. The item they believe is the most likely to be the signal item receives rank 1, the item they believe is the second most likely to be the signal item receives rank 2, etc. Thus, the results from a multi-item ranking task is the probability distribution with which the signal item receives each of the possible ranks. To perform the critical test, the distribution of ranking responses needs to be transformed into the corresponding hazard function, representing the probability that the signal item would appear at the current rank given that it had not appeared before. Chechile and Dunn provide a proof on the shape of the hazard function under a threshold representation. A threshold process would imply that participants either detect the signal (i.e., evidence exceeds the threshold), or else, they guess among the remaining items. For the hazard function to be feasibly generated by such a threshold process, it must then be *monotonic* after rank 1, with the probability of the signal appearing at any rank after rank 1 being only a guess among the remaining ranks (increasing as the remaining ranks decrease).

Our application of Chechile and Dunn's (2021) critical test for threshold representation will not use an explicit ranking data, but will use the same m-AFC data used for the test of RSR. We therefore first show that the hazard function can be derived indirectly from m-AFC accuracy data. We then carry out a critical test as to whether this hazard function is monotone after rank 1. If this test is passed, then models assuming that there is a threshold for evidence-strength are admissible, meaning that it remains feasible that global motion is an all-or-nothing process. If the test fails, then all models assuming a threshold decision rule are inadmissible, regardless of parametric instantiations. This would provide evidence that global motion detection is governed by a continuous-strength model.

Before presenting results from our own data, we show that the critical test of threshold representation is diagnostic and can provide evidence against threshold representations. In particular, we show that it can do so when using the hazard function derived from m-AFC data (as we also plan to do). For this we apply the critical test to the recognition memory data reported by Kellen et al. (2021). Recall that there exists independent evidence against threshold representation for this data, thus we would expect this test to fail. Figure 2 shows the hazard functions derived from m-AFC data from Kellen et al.'s Experiments 1 and 2, as well as fitted monotonic values (for technical details see below). The large and significant G^2 values indicate that the empirical hazard deviates from the nearest monotone hazard, and therefore the data fails the critical test and provides evidence against a threshold representation. The deviations from monotonicity, between ranks 2 and 4, match the position where continuous SDT models are expected to fail (Chechile & Dunn, 2021).

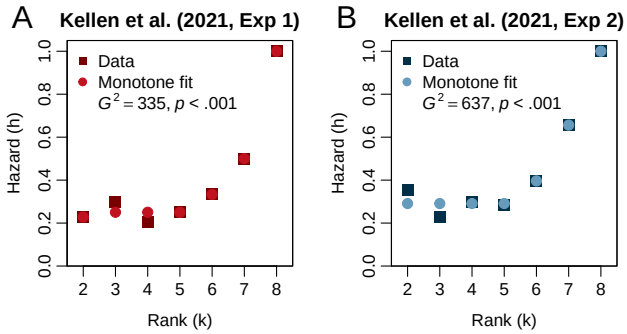

Figure 2: Non-monotonic hazard functions derived from m-AFC data that defy threshold representation (Kellen et al., 2021, (A) Experiment 1 and (B) Experiment 2).

The Present Study

The aim of the present study was to test both RSR and threshold representation in global motion detection. To this end, we performed an m-AFC experiment with $m = 2$ to $m = 8$ in steps of 1. RSR imposes constraints on the accuracy across m and can be directly tested from the m-AFC data. Threshold representation imposes constraints on the hazard function of the ranking task, which we will derive from the m-AFC data. Finally, we also reconstruct the Yes-No receiver operating characteristic (ROC) function from the m-AFC data. Threshold models are characterised by linear ROCs. Thus, if we find evidence for a threshold representation, we should also find linear ROCs.

Method

Participants

101 participants (49 males, mean age = 45.7) with normal or corrected vision were recruited online via Prolific Academic and reimbursed at a rate of £9/hour. Participants were briefed on the procedure and provided informed consent. The study was approved by the Ethics Committee of the UCL Research Department of Experimental Psychology, ID No:

EP/2021/005. All participants performed the study online on their own personal computers.

Design

A within-subject design was used, varying choice set size m across trials from $m = 2$ to $m = 8$ in step size of 1 (i.e., seven different levels of m). Participants completed 10 trials per m in a random order for a total of 70 trials. The experiment was programmed in JavaScript using `JsPsych` and `jspsych-rdk` (Rajananda et al., 2018).

Stimuli

Stimuli were shown on a black screen populated with $m \in \{2, \ldots, 8\}$ apertures (outlined circles), randomly placed in m of 8 placeholders in a circle around the centre of the screen. An example of a 6-AFC trial is shown in Figure 1B. Each aperture contained 100 white dots, whose position changed in every frame ($\approx 17ms$). In every trial, one aperture – the signal – contained some dots moving in a single direction, along with dots moving in randomly. The other $m - 1$ apertures – the noise – contained only dots moving randomly. Participants' task was to select the signal aperture by clicking on it.

In the signal aperture, the direction of global motion was randomly chosen at each trial from the set of angles which are multiples of 45 degrees. At each frame 12% of the dots were randomly chosen to be moving in the direction of global motion (i.e., there were no pre-designated coherent dots; they were assigned randomly at each frame). Coherence was set to 12% after pilot testing such that the task was challenging even without a time-out, but still doable. In any frame, dots not designated to move in the direction of global motion were moving randomly. Random movement followed a random walk, meaning the dot would take one step in a random direction during that frame.

In noise apertures there was no coherent motion. All dots were in random motion for all frames, moving according to random walks. In all apertures, any dot that reached the aperture boundary disappeared and a new dot was generated at a random location. All dots moved at a speed of 120 pixels/second.

Procedure

Having read the instructions, each participant completed 7 practice trials, 1 for each m, presented in a set order. Only during the practice trials did participants receive correct/incorrect feedback on their responses. The practice trials were followed by a brief instruction refresher, which was followed by the main task. In the main task, participants worked on 70 trials, 10 for each m, which were presented in a random order, with no feedback. There was no time-out, so participants only proceeded to the next trial after mouse-clicking on an aperture. After a response, a black screen appeared for 500 ms between trials. After completing all trials, participants were debriefed and given performance feedback. The experiment required full-screen mode throughout.

Results

Three participants were excluded for reporting that they had not taken the task seriously. All results below are based on the remaining dataset of 98 participants, offering 980 trials per m. Kellen et al. (2021) showed that 500 trials per m-AFC condition corresponds to a statistical power of 95% for detecting violations of random-scale representation (RSR), assuming responses come from a uniform distribution, so the experiment is well-powered to detect violations of RSR.

As a robustness check, participants whose performance did not significantly exceed chance were also excluded. After all exclusions, 73 participants remained (730 trials per m). All results reported below replicated for this reduced sample.

For data and analysis scripts, see https://osf.io/uvc8p/.

Random-Scale Representation Critical Test

We first assessed whether the global motion detection data satisfy random-scale representation (RSR). To this end, we began by aggregating the m-AFC data to get accuracy proportions $P_S^{\langle m \rangle}$ for each $m \in \{2, \ldots, 8\}$. We constructed a matrix containing the Block-Marschak Inequalities and the likelihood-ratio monotonicity constraints, defining the constrained outcome space in which RSR is satisfied. We then projected the observed accuracy proportions $P_S^{\langle m \rangle}$ onto this constrained space, using quadratic programming (via the `quadprogpp` R-package) to minimise the squared distance between the fitted values $\hat{P}_S^{\langle m \rangle}$ and the observed values $P_S^{\langle m \rangle}$ (Beale, 1959).

Figure 3A shows both the empirical m-AFC accuracies $P_S^{\langle m \rangle}$ and the RSR-fitted values $\hat{P}_S^{\langle m \rangle}$. It is clear from this figure that the empirical data fit the RSR constraints very well, providing strong evidence for RSR in global motion detection.

To assess the goodness-of-fit, we compared the expected counts under the RSR-fitted values to the observed counts, via the G^2 statistic. To assess whether the divergence between observed and expected counts was significant, we employed a double bootstrap procedure to generate a null distribution of G^2 under the assumption that RSR holds, propagating sampling variability through the constraint fitting (for details see Kellen et al., 2021).

The observed misfit is small with $G^2 = 5.78$, and non-significant, with double-bootstrapped $p = .230$ (10,000 iterations). The experiment therefore passes the critical test, which provides strong evidence that a RSR must underlie judgements in the random dot motion task.

Threshold Representation Critical Test

Next, we assessed whether the global motion detection data satisfy Chechile and Dunn's (2021) test for threshold representation. To this end, we first derived rank probabilities and from these the implied hazard function. From the RSR-fitted m-AFC accuracy values, $\hat{P}_i^{\langle m \rangle}$, we derived the implied rank probabilities $R_i^{\langle m \rangle}$. $R_i^{\langle m \rangle}$ is the probability that the signal is at rank i out of m stimuli, or equivalently, the probability that the signal is exceeded by $i - 1$ of the $m - 1$ noise items. Assuming

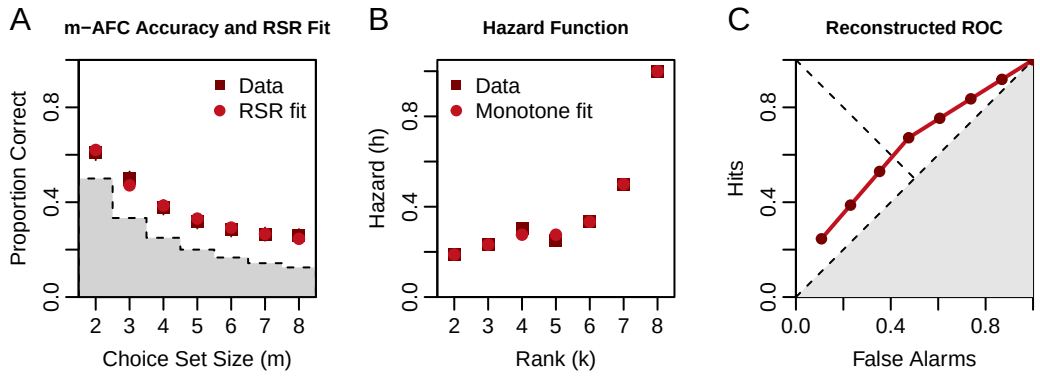

Figure 3: Results. A: Empirical and RSR-fitted m-AFC accuracy values per m. The grey region shows chance performance. B: Empirical hazard function derived from RSR-fitted values, and closest-fitting monotone hazard function, over possible signal ranks. C: Reconstructed Yes-No ROC curve showing implied hits to false alarms. Shape appears piece-wise linear.

that the latent strengths of the m alternatives are independent, we represented $R_i^{\langle m \rangle}$ as a linear combination of m-AFC probabilities with a binomial coefficient. The closed-form mapping between $\hat{P}_i^{\langle m \rangle}$ and $R_i^{\langle m \rangle}$ was obtained as:

$$R_i^{\langle m \rangle} = \binom{m-1}{i-1} \sum_{j=m-i+1}^{m} \binom{i-1}{j-(m-i+1)} (-1)^{j-(m-i+1)} \hat{P}_S^{\langle j \rangle}.$$

Focusing on $m = 8$, which describes the full rank distribution $R_1, \ldots, R_8$, we converted the signal ranking probabilities into a hazard function defined as,

$$h_k = \frac{R_k}{\sum_{j \geq k} R_j}, \qquad k = 1, \ldots, 8.$$

We then fit the empirical (derived) hazard vector $(h_2, \ldots, h_8)$ via quadratic programming to the monotonicity and boundary constraints $(h_{k+1} \geq h_k, 0 \leq h_k \leq 1)$, yielding the closest monotone hazard function, $\hat{h}_k$.

Figure 3B shows the hazard function derived from empirical data h_k and the closest fitting monotone hazard $\hat{h}_k$. The data shows one apparent violation of monotonicity at ranks 4 and 5. To answer the question whether the critical test for threshold representation holds, we therefore need to know whether this violation is statistically significant or not.

To evaluate the degree of misfit from monotonicity, we followed the same general procedure as for the critical test for RSR. One complication in this process was that it required us to invert our derivation of the hazard function to recover accuracy proportions implied by the fitted hazard values, $\hat{h}_k$. The resulting derived accuracy proportions entailed the expected counts under the monotone hazard function, and we compared these to the observed counts, again via the G^2 statistic. We again assessed statistical significance using a double bootstrap procedure to approximate the null distribution of G^2 under the monotone-hazard assumptions.

Again, the observed misfit was small, $G^2 = 5.90$, double-bootstrapped $p = .281$ (10,000 iterations). Recall that for recognition memory data with an approximately equal sample size (Figure 2) we obtained a significant p-value using exactly the same procedure. Thus, the non-significant goodness-of-fit test provides strong evidence that global motion data can be produced under a monotonic hazard function, where detection always occurs above a threshold, and mere guessing occurs below it (Chechile & Dunn, 2021). In sum, the global motion data passes the critical test for threshold representation.

Yes-No ROC Reconstruction

Finally, we reconstructed the Yes-No ROC function using rank probabilities, $R_i^{\langle m \rangle}$, derived from the fitted accuracy values as shown above (for details see He et al., 2026; Kellen et al., 2021). The reconstructed ROC, shown in Figure 3C, can be compared to ROCs expected under different models to provide evidence for their suitability. Here we compare it to the ROCs predicted by equal-variance SDT and one-high threshold model (we focus on the one-high threshold model instead of the two-high threshold model, because of the inability to detect a noise item as noise in global motion detection; Macmillan & Creelman, 2005).

The reconstructed Yes-No ROC function (Figure 3C) appears piece-wise linear and asymmetric. This provides some evidence against the equal-variance Gaussian SDT model, which would predict a curved, convex, and symmetric ROC function. On the other hand, linear ROCs are congruent with threshold representations, and thus the ROC provides some evidence in support of global motion detection being reliant on evidence thresholds. However, the reconstructed ROC function does not match the predicted ROC shape of the one-high threshold model, which is a straight line starting at (1,1) (Macmillan & Creelman, 2005). While the critical test of threshold representation has shown that global motion detection can feasibly be modelled by a threshold model and the reconstructed ROC curve is piece-wise linear consistent with such models, the reconstructed ROC is inconcsistent with the most obvious threshold model candidate. Whether there is another threshold model variant that can account for the observed ROC shape is an open question for future research.

Discussion

The aim of the current study was to test whether global motion detection empirically satisfies random-scale representation (RSR) and threshold representation. To this end, we carried out an m-AFC random dot motion experiment, in which participants were tasked with selecting the aperture containing coherent motion out of m alternatives, for $m = \{2 \dots 8\}$ in steps of 1. To test the RSR assumption, we fit the m-AFC accuracy data to the Block Marschak Inequalities (Block & Marschak, 1974) and likelihood-ratio monotonicity constraints. To test threshold representation, we derived a hazard function from the RSR-fitted m-AFC values, and fit this hazard to a monotonicity constraint (Chechile & Dunn, 2021). We found strong evidence that global motion detection satisfies RSR and evidence consistent with threshold representation. Additionally, we reconstructed a Yes-No ROC curve from the m-AFC data, using implied signal rankings. The ROC curve was piece-wise linear, providing some evidence against the equal-variance Gaussian assumption and for a threshold representation. The reconstructed ROC did not, however, match the ROC predicted by the one-high threshold model, and so an open question remains as to which parametric model can adequately describe global motion detection.

The strongest evidence in our work is for global motion detection requiring models with RSR, which includes SDT and threshold models. Observed m-AFC accuracy obeyed the RSR constraints, which necessarily implies that a model with RSR underlies these judgements (Falmagne, 1978). For our critical test, RSR entails that participants maintain the same decision strategy across m. Had they changed their decision strategy across m – for example, only considering a subset of items for larger m – we would expect a violation of RSR (He et al., 2026; Kellen et al., 2021). Thus, we can say that it is very unlikely that participants changed their decision strategies across m in our task.

One of the assumptions of the SDT framework is that in the m-AFC task participants use the "max" decision strategy, selecting the aperture with the maximum evidence strength. Our results shown here are not sufficient to test whether participants indeed used the "max" strategy or a different strategy that is in principle also compatible with RSR, such as a serial search strategy. One way to test the "max" decision strategy is through tests based on the generalised area theorem (Iverson & Bamber, 1997), which only holds under the "max" decision strategy. For example, combining an m-AFC task with a Yes/No task and seeing whether the Yes/No Accuracy is on top of the m-AFC implied ROC curve (e.g., Kellen et al., 2021, Exp. 2). In visual working memory, He et al. (2026, Exp. 1b) have provided further evidence that participants follow the "max" decision strategy.

We have also provided support for M. L. Green and Pratte (2022)'s claim that global motion detection is an all-or-nothing process, both because global motion performance passed the critical test for threshold representation and because the reconstructed ROC curve is piece-wise linear rather than curved. That said, the evidence for threshold representation is less strong than the evidence for RSR. The first reason for this difference in evidence strength comes from the formal properties of both critical tests. As proved by Falmagne (1978), if a set of decisions obeys the constraints of RSR, this means a model with RSR *must* underlie these decisions. On the other hand, the critical test of Chechile and Dunn (2021) is formally a test *against* threshold representation. If the test fails, we know that a model with threshold representation cannot be responsible for the observed data. However, if the test passes we do not have any certainty; a threshold model is clearly possible but not necessary. In fact, Chechile and Dunn have shown that in the case of low discriminability, even equal-variance Gaussian SDT models can produce a monotone hazard function. Thus, providing compelling evidence for threshold representation requires showing a monotone hazard function under various levels of discriminability.

The second reason for the difference in evidence strength is that the critical test for threshold representation was performed indirectly. It relied on deriving rank probabilities from m-AFC data assuming RSR, instead of collecting ranking data directly. Showing that the critical test of Chechile and Dunn can be done from only m-AFC data, not requiring an explicit ranking task, is a genuine contribution of the present work. However, this comes at the cost of a more direct test. It is possible that by transforming necessarily noisy m-AFC data into ranking probabilities, we affect the monotonicity of the hazard function. Of course, the results displayed in Figure 2 that show a violation of monotonicity for the derived hazards from m-AFC data when this is expected provide some assurances of the validity of our procedure. Nevertheless, future work testing threshold representation in global motion detection would benefit from a direct ranking task.

There is an emerging body of work in the memory literature that performs critical tests of the assumptions underlying choice models in order to validate whole families of models that are commonly used in this domain (Dunn et al., 2025; He et al., 2026; Kellen et al., 2021; McCormick & Semmler, 2023; Meyer-Grant et al., in press). The current paper extends these critical tests to the perceptual domain, particularly global motion detection. Global motion detection is, however, only one example of a perceptual task where d' is used without testing its underlying assumption of equal-variance Gaussian evidence distributions. The critical tests we have deployed on global motion detection data are in theory suitable for any domain permitting an m-AFC task with variable m. Ideally, the stimuli should be constructed in a way that all m stimuli are visible at the same time to limit the role of memory in performance. Whether a task in which stimuli are only shown for a brief time, a common procedure in many perceptual decision-making tasks (e.g., Bang et al., 2018; Kvam & Pleskac, 2016; Shekhar & Rahnev, 2024), would still permit performing these critical tests is a key question for future research, as this could allow for testing of modelling assumptions across a wide range of perceptual domains.

Acknowledgments

We thank David Kellen and Yiou He for helpful discussions.

References

Bang, J. W., Shekhar, M., & Rahnev, D. (2018). Sensory Noise Increases Metacognitive Efficiency. *Journal of Experimental Psychology: General*, *148*(3), 437–452. https://doi.org/http://dx.doi.org/10.1037/xge0000511

Barlow, H., & Tripathy, S. P. (1997). Correspondence Noise and Signal Pooling in the Detection of Coherent Visual Motion. *The Journal of Neuroscience*, *17*(20), 7954–7966. https://doi.org/10.1523/JNEUROSCI.17-20-07954.1997

Beale, E. M. L. (1959). On quadratic proramming. *Naval Research Logistics Quarterly*, *6*(3), 227–243. https://doi.org/10.1002/nav.3800060305

Block, H. D., & Marschak, J. (1974). Random Orderings and Stochastic Theories of Responses (1960). In *Economic Information, Decision, and Prediction* (pp. 172–217). Springer Netherlands. https://doi.org/10.1007/978-94-010-9276-0_8

Britten, K., Shadlen, M., Newsome, W., & Movshon, J. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience*, *12*(12), 4745–4765. https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992

Chechile, R. A., & Dunn, J. C. (2021). Critical tests of the two high-threshold model of recognition via analyses of hazard functions. *Journal of Mathematical Psychology*, *105*, 102600. https://doi.org/10.1016/j.jmp.2021.102600

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. *Cognition*, *207*, 104522. https://doi.org/10.1016/j.cognition.2020.104522

Desender, K., Vermeylen, L., & Verguts, T. (2022). Dynamic influences on static measures of metacognition. *Nature Communications*, *13*(1), 4208. https://doi.org/10.1038/s41467-022-31727-0

Dunn, J. C., Anderson, L. M., & Stephens, R. (2025, November 17). *A shift representable signal detection model of recognition memory.* https://doi.org/10.31234/osf.io/6mv75_v1

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*(1), 52–72. https://doi.org/10.1016/0022-2496(78)90048-2

Green, D. M. (2020). A homily on signal detection theory. *The Journal of the Acoustical Society of America*, *148*(1), 222–225. https://doi.org/10.1121/10.0001525

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.

Green, M. L., & Pratte, M. S. (2022). Local motion pooling is continuous, global motion perception is discrete. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(1), 52–63. https://doi.org/10.1037/xhp0000971

Green, M. L., & Pratte, M. S. (2026). Perceptual confidence has near perfect access to the existence of discrete representations, but only weak access to precision. *Cognition*, *267*, 106347. https://doi.org/10.1016/j.cognition.2025.106347

He, Y., Kellen, D., & Singmann, H. (2026, January 28). *Examining the Signal-Detection Account of Visual Working Memory*. w287y_v3. https://doi.org/10.31219/osf.io/w287y_v3

Iverson, G., & Bamber, D. (1997). The generalized area theorem in signal detection theory. In *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 301–318). Lawrence Erlbaum & Associates.

Kellen, D., & Klauer, K. C. (2018). Elementary Signal Detection and Threshold Theory. In J. T. Wixted (Ed.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–39). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119170174.epcn505

Kellen, D., Winiger, S., Dunn, J. C., & Singmann, H. (2021). Testing the foundations of signal detection theory in recognition memory. *Psychological Review*, *128*(6), 1022–1050. https://doi.org/10.1037/rev0000288

Kvam, P. D., & Pleskac, T. J. (2016). Strength and weight: The determinants of choice and confidence. *Cognition*, *152*, 170–180. https://doi.org/10.1016/j.cognition.2016.04.008

Lappin, J. S., & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Research*, *16*(2), 161–168. https://doi.org/10.1016/0042-6989(76)90093-6

Liang, J., Zhou, Y., & Liu, Z. (2016). Examining the standard model of signal detection theory in motion discrimination. *Journal of Vision*, *16*(7), 9. https://doi.org/10.1167/16.7.9

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory : A user's guide*. Lawrence Erlbaum associates.

McCormick, K. M., & Semmler, C. (2023, July 7). *Letting go of the Grail: Falsifying the theory of 'true' eyewitness identifications*. https://doi.org/10.31234/osf.io/dszqw

Meyer-Grant, C. G., Kellen, D., Harding, S., & Singmann, H. (in press). Extreme-Value Signal Detection Theory for Recognition Memory: The Parametric Road Not Taken. *Psychological Review*. https://osf.io/qhrfj

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54. https://doi.org/10.1038/341052a0

Rajananda, S., Lau, H., & Odegaard, B. (2018). A Random-Dot Kinematogram for Web-Based Vision Research. *Journal of Open Research Software*, *6*(1), 6. https://doi.org/10.5334/jors.194

Schütz, A. C., Braun, D. I., Movshon, J. A., & Gegenfurtner, K. R. (2010). Does the noise matter? Effects of different kinematogram types on smooth pursuit eye movements and perception. *Journal of Vision*, *10*(13), 26. https://doi.org/10.1167/10.13.26

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Sciences*, *93*(2), 628–633. https://doi.org/10.1073/pnas.93.2.628

Shekhar, M., & Rahnev, D. (2024). How Do Humans Give Confidence? A Comprehensive Comparisonof Process Models of Perceptual Metacognition. *Journal of Experimental Psychology: General*, *153*(3), 656–688. https://doi.org/https://doi.org/10.1037/xge0001524

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*(1), 55–62. https://doi.org/10.1016/0042-6989(84)90144-5