# The metaperceptual function: Exploring dissociations between confidence and task performance with type 2 psychometric curves

Brian Maniscalco[1,2*†], Olenka Graham Castaneda[2*], Brian Odegaard[3], Jorge Morales[4], Sivananda Rajananda[2], Megan A. K. Peters[1,2]

## Highlights:

- Under certain experimental conditions, confidence and task performance dissociate
- This effect provides an important tool for studying confidence and awareness
- We introduce metaperceptual (type 2 psychometric) curves to study such dissociations
- We characterize how such dissociations behave across a wide range of conditions
- These findings can inform the design of future research on confidence and awareness

# The metaperceptual function: Exploring dissociations between confidence and task performance with type 2 psychometric curves

**Authors:**

Brian Maniscalco[1,2*†], Olenka Graham Castaneda[2*], Brian Odegaard[3], Jorge Morales[4], Sivananda Rajananda[2], Megan A. K. Peters[1,2]

**Affiliations:**

[1] Department of Cognitive Sciences, University of California Irvine, Irvine, CA 92697
[2] Department of Bioengineering, University of California Riverside, Riverside, CA 92521
[3] Department of Psychology, University of Florida, Gainesville, FL, 32608
[4] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21218

* These authors contributed equally.
† Correspondence should be addressed to:
Brian Maniscalco (bmanisca@uci.edu)
Department of Cognitive Sciences
University of California, Irvine
2201 Social & Behavioral Sciences Gateway Building
Irvine, CA 92697-5100

# Abstract

Confidence can dissociate from perceptual accuracy, suggesting distinct computational and neural processes underlie these psychological functions. Recent investigations have therefore sought to experimentally isolate metacognitive processes by creating conditions where perceptual sensitivity is matched but confidence differs ("matched-performance / different-confidence"; MPDC). Despite these endeavors' success, much remains unknown about MPDC effects and how to best harness them in experimental settings. Here we developed a principled approach to comprehensively characterizing MPDC effects through analyzing metaperceptual (i.e., type 2 psychometric) functions relating objective performance to subjective confidence across widely varying performance levels and experimental manipulations. We found that MPDC effect magnitude depends on stimulus properties, observers' sensitivity level, and critically on trial type order (blocked or interleaved across stimulus property variations). Our findings provide the first comprehensive exploration of MPDC effects, offer a prescriptive guide to metaperceptual analysis, and suggest optimal experimental paradigms for experimentally isolating metacognition and awareness in future studies.

**Keywords:** confidence; metacognition; performance matching; signal detection theory; psychometric function

# The metaperceptual function: Exploring dissociations between confidence and task performance with type 2 psychometric curves

## 1. Introduction

Our perceptual decisions are typically accompanied by a subjective feeling of confidence: “I’m sure I saw Austin at the store,” “I can’t tell at what speed the car is coming at me,” “This chip of paint looks identical to this other one, but I’m not totally sure.” When making simple perceptual decisions like these, confidence often tracks how accurate we are in a particular task: accurate decisions tend to produce higher confidence, and inaccurate decisions tend to produce lower confidence. For example, in a laboratory setting, task difficulty often correlates with accuracy and confidence (the harder the task, the less accurate and the less confident participants tend to be) (Baranski & Petrusic, 1994). However, task performance and subjective confidence can dissociate. These dissociations have been observed after brain lesion (Azzopardi & Cowey, 1997; Del Cul et al., 2009; Fleming et al., 2010; Weiskrantz, 1986), experimental manipulation (Cortese et al., 2016; Koizumi et al., 2015; Lau & Passingham, 2006; Maniscalco et al., 2016; Odegaard et al., 2018; Peters, Fesi, et al., 2017; Rahnev, Maniscalco, et al., 2012; Rollwage et al., 2020; Rounis et al., 2010; Samaha et al., 2016; Stolyarova et al., 2019), and spontaneous fluctuations in neural signals (Rahnev, Bahdo, et al., 2012; Samaha et al., 2017), and there are also natural individual differences in neurotypical individuals (Fleming et al., 2010). Importantly, this dissociation between task performance and subjective confidence can be leveraged to further our understanding of the behavioral, computational, and neural profile of subjective confidence and subjective feelings of awareness (Lau & Passingham, 2006; Miyoshi & Lau, 2020; Morales et al., 2019; Odegaard et al., 2018; Peters et al., 2016; Peters, Fesi, et al., 2017; Peters, Thesen, et al., 2017; Stolyarova et al., 2019).

In the laboratory, specific alterations of stimuli in visual psychophysical experiments can produce pairs of conditions which yield matched performance and different confidence (Koizumi et al., 2015; Lau & Passingham, 2006; Odegaard et al., 2018; Rollwage et al., 2020; Samaha et al., 2016; Stolyarova et al., 2019; Zylberberg et al., 2012). By keeping performance constant while obtaining diverging subjective reports — sometimes referred to as a “matched-performance / different confidence” (MPDC) effect — subjective confidence can be isolated and performance neutralized as a potential confound (Lau, 2008; Morales et al., 2015, 2019; Peters et al., 2016). MPDC effects have been shown for several types of stimuli (including dot motion patterns and visual gratings (Koizumi et al., 2015; Odegaard et al., 2018; Rollwage et al., 2020)) and induction methods (including manipulating stimulus variability and levels of positive and negative evidence (see (Morales et al., 2019) for a review)), and are robust and replicable (Samaha et al., 2016). Moreover, MPDC effects have been predicted by general

principles from signal detection theory, which provides a useful framework for understanding why the effects occur (see (Morales et al., 2019) for further details).

However, much is still unknown about this phenomenon. The general signal detection-theoretic account of matched-performance different-confidence predicts that the effect should emerge over a broad range of task performance (see e.g. Figure S7), but this has not been systematically tested. For example, across a range of performance levels (from chance-level to ceiling-level performance), are targeted stimulus manipulations equally successful in producing the effect? Or is there something akin to a "sweet-spot" in a specific performance range where the dissociation emerges? A full psychophysical characterization of this kind of dissociation is currently lacking, and much is still unknown about the conditions under which "matched-performance / different-confidence" (MPDC) effects are possible.

Other design choices may also influence MPDC effects: when different conditions are used in a single task, trial types can either be randomized or grouped in a block design. Do these types of considerations influence the prevalence of the effect? Randomly interleaving conditions can help facilitate a constant decision strategy, since human participants have difficulty dynamically adjusting response criteria from trial to trial in response to frequent changes to stimulus characteristics (Brown & Steyvers, 2005; Gorea & Sagi, 2000). Thus, whether interleaved or block designs influence MPDC effects remains to be explored.

Here, we present results from an experiment where we collected participants' perceptual decisions and confidence ratings in a simple random dot kinetogram (RDK) task. Participants viewed whole-screen random dot motion presented continuously throughout each block of trials. On every trial, a circular region of the screen to the left or right of fixation transiently exhibited coherent downward motion. Participants indicated whether the coherent motion occurred on the left or right side of fixation, and then rated confidence in the accuracy of their decision. The stimuli were presented at seven levels of coherence and three levels of dot density, creating 21 trial types. Following previous demonstrations (Koizumi et al., 2015; Odegaard et al., 2018; Rollwage et al., 2020; Samaha et al., 2016; Stolyarova et al., 2019), we expected higher dot density conditions to yield higher confidence, even when task performance was similar. Density levels were either "Blocked" (constant dot density within a block of trials) or "Interleaved" (dot density changing across trials within a block), providing an opportunity to evaluate the effects of slowly and rapidly changing stimulus features on participants' confidence rating behavior. We also present a comprehensive exploration of best practices for quantifying the type 2 psychometric function relating type 1 task performance to type 2 metacognitive judgments, which we term the *metaperceptual* function for short (Figure 1). Together, the results of these experiments and our analytic approach provide for the first time a systematic characterization of MPDC phenomena over full metaperceptual functions, and establish a guide for future efforts quantifying these effects in service of optimizing experimental and analytic approaches to reveal the neural and computational correlates of metacognition.

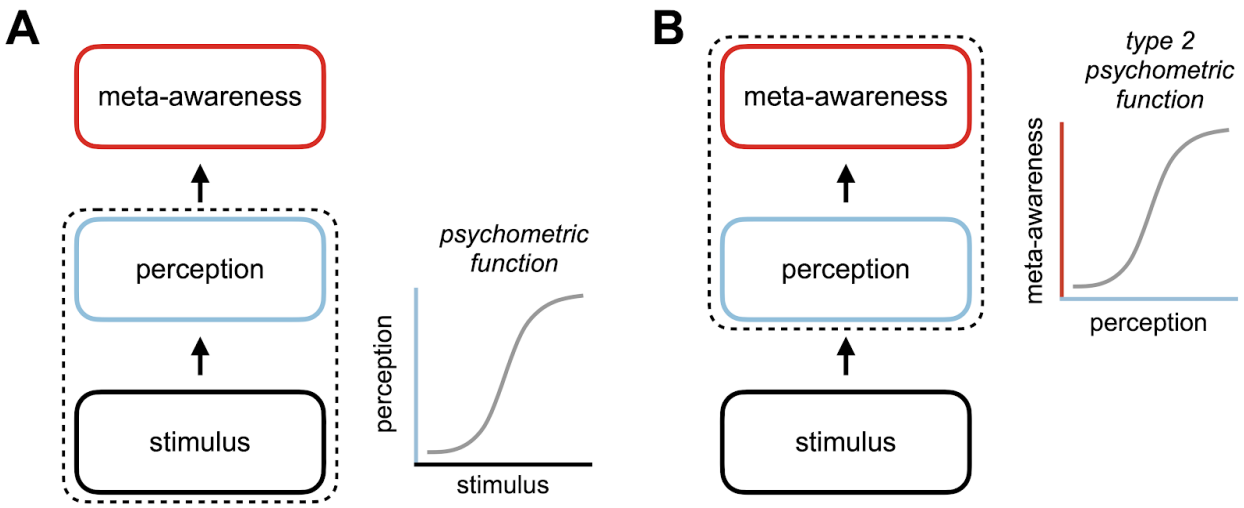


**Figure 1. The type 2 psychometric function, or *metaperceptual* function.** (A) Conventional psychometric functions characterize the relationship between objective stimulus features (e.g., contrast, motion coherence, etc.) and an observer's perception of that stimulus (e.g., probability of detecting the stimulus, probability of accurately discriminating a stimulus feature, etc.) by plotting perceptual performance against stimulus properties. (B) By way of analogy, here we introduce the concept of a *type 2 psychometric function*, which characterizes the relationship between perceptual task performance and "meta-awareness" (i.e., an observer's experiences *of* or judgments *about* perception). Possible measures of meta-awareness are, for example, ratings of subjective perceptual experience, ratings of confidence about perceptual decision accuracy, or the accuracy with which such ratings predict perceptual performance. Relative to the axes of a conventional psychometric function, the axes of a type 2 psychometric function are both shifted one level upwards on the hierarchy of {stimulus, perception, meta-awareness}, such that meta-awareness is plotted as a function of perceptual performance. We also coin the term *metaperceptual function* to be synonymous with "type 2 psychometric function," where "metaperceptual" functions analogously to the term "psychophysical." Just as the roots of the word "psychophysical" connote "relationship of perception (psycho) to stimulus (physical)," so the roots of "metaperceptual" connote "relationship of meta-awareness to perception."

## 2. Methods

### 2.1. Participants

27 University of California Riverside students (19 female, 8 male, 26 right-handed, mean age = 20.6 (SD = 3.1)) provided written informed consent to participate in the main study. All participants had normal or corrected-to-normal vision and normal or corrected-to-normal hearing, and were compensated at a rate of $10/hour for their participation. All study procedures were approved by the University of California Riverside Institutional Review Board.

Prior to the main group-level analysis, data from individual participants were inspected for quality. Data from six participants were excluded from the main analysis due to having performance at or near chance levels across all motion coherence levels (n=3), having

completely flat (n=1) or excessively noisy (n=1) confidence vs d' curves, and using a single confidence rating on almost all trials (n=1). Therefore, 21 participants were included in the main analyses reported below.

2.2. Stimulus & equipment

All stimuli were presented on a CRT monitor (NEC MultiSync FE2111SB-BK, width 39.6 cm, height 29.7 cm) with refresh rate 75 Hz. A random dot kinematogram (RDK) filling the entire screen (width x height = 43.2 x 33.1 degrees of visual angle (deg)) was presented continuously throughout every block of trials. Dots were black on a white background, with dot size = 0.1 deg, speed = 6 deg/sec, and lifetime = 67 ms (5 frames). When a dot's lifetime expired, it was removed from the screen and replaced with a new dot having a full lifetime and randomly determined location and motion direction. At the start of each block, dots were initialized with uniformly distributed "age," such that on every frame refresh of the screen, one-fifth of the dots expired and were respawned. Dots that moved outside the bounds of the screen continued their motion trajectory from the opposite side of the screen.

Dot density took on one of three possible values (Low = 1 dot/deg$^2$, Medium = 3 dots/deg$^2$, High = 9 dots/deg$^2$), and was varied either across blocks (Blocked condition) or across trials within a block (Interleaved condition). When dot density decreased from trial N to trial N+1, a randomly selected portion of the dots were deleted in order to achieve the appropriate density. When dot density increased, an appropriate number of new dots were spawned with uniformly distributed age and randomly selected location and motion direction.

A fixation cross (width = 0.35 deg) was presented in the center of the screen. Color of the fixation cross changed depending on trial state (see below). Participants were instructed to maintain fixation on the fixation cross throughout each block. To prevent dots from visually interfering with the fixation cross, any dots whose locations fell inside a small circular region in the center of the screen (diameter = 2 deg) were not displayed.

The critical stimulus event occurring on every trial was the occurrence of 533 ms of coherent downward motion in a circular region of the screen (diameter = 8 deg) whose center was located 7 deg to the left or right of fixation, which we will call the "region of coherence." Motion coherence was drawn from one of seven possible values spaced evenly between 10% and 80%, i.e. [10, 21.67, 33.33, 45, 56.67, 68.33, 80]%.

Coherent motion was created by assigning downward motion to all dots spawned with initial locations falling within the region of coherence with probability p(motion coherence) for a period lasting 493 ms (37 frames). Thus, onset and offset of motion coherence was temporally smoothed due to being yoked to dot respawning, which occurred for one-fifth of the dots on every frame. In total, motion coherence linearly ramped up during the first 53 ms (4 frames) of motion coherence, remained at full motion coherence for the next 427 ms (32 frames), and then linearly ramped down during the final 53 ms (4 frames). Additionally, since motion direction for

every dot was constant throughout its lifetime, there were no sharp perceptual edges around the perimeter of the region of coherence due to abrupt changes in dot motion direction as dots entered and exited the region.

### 2.3. Procedure

Participants sat approximately 50 cm from the screen with their chins in a chinrest. Each trial began with presentation of full-field random dot motion for a pre-stimulus period lasting 1 - 3 s. Pre-stimulus duration was drawn randomly from an exponential distribution on each trial such that the hazard rate was roughly held constant; this meant that during the pre-stimulus period, the amount of time elapsed so far was made to be uninformative about whether the target stimulus was about to occur. During this period the fixation cross was red in order to cue the subject to be ready to detect impending coherent motion. Subsequently, the fixation cross turned black and coherent downward motion appeared in one of the two circular regions of coherence (533 ms). The region of coherence was equally likely to appear on either the left or right side of fixation.

After stimulus offset, participants were given three seconds to report the side in which they saw the downward movement (by pressing the 1 or 2 key) and how confident they were in their judgement on a scale of 1 to 4 (using the 7 8 9 0 keys). On trials where participants could not clearly make out the location of coherent motion, they were encouraged to enter a response anyway by making a random guess. To provide feedback on registry of keyboard input, the fixation cross turned gray after entry of the left / right decision and disappeared after entry of confidence. The full 3 s of the response period played out even on trials where participants entered their perceptual decision and confidence rating prior to the expiration of the 3 s time limit. A schematic of trial structure is shown in Figure 2A.

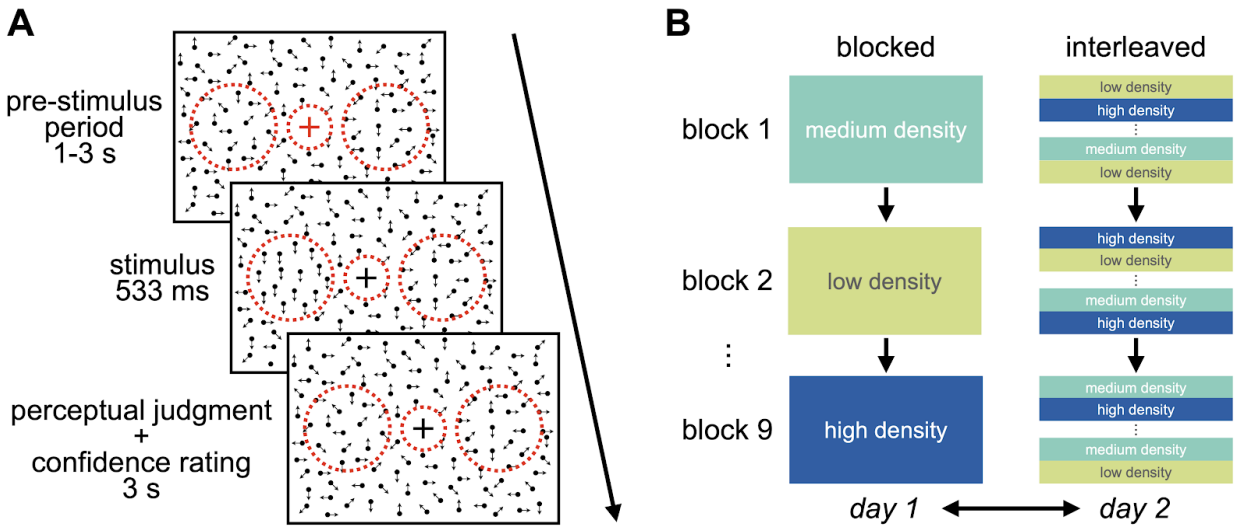


**Figure 2. Behavioral task procedures.** (A) Each trial began with a 1-3 s pre-stimulus period, during which full-field random dot motion was shown (black arrows illustrate dot motion direction). Subsequently, within one of two circular regions of the screen (indicated here by the red circles to the left and right of fixation—red circles were shown to participants only during preliminary practice trials but not during experimental trials), coherent downward dot-motion occurred, followed by a response period in which participants indicated on which side they saw the coherent motion and their decision confidence. The central red circle indicates an area around the fixation cross where no dots were presented; this red circle was not shown to participants and is used here for illustration purposes. (B) Participants underwent two trial-order conditions, Blocked and Interleaved, on two different days of testing. In the Blocked condition, dot density was constant across trials within a given block, whereas in the Interleaved condition, dot density varied randomly across trials. Blocked versus Interleaved days and order of density blocks was counterbalanced across all participants.

## 2.4. Blocked versus Interleaved design

Participants underwent two trial-order conditions in which dot density was either presented pseudorandomly across trials in an *Interleaved* design, or was *Blocked* by dot density. In the Interleaved type block, the density level on each trial was pseudorandomly drawn from any of the three density levels (Low, Medium, or High); in the Blocked condition, all trials within a block had the same density. In both conditions, within each block of trials all coherence levels were presented in pseudorandom order.

The order of the Blocked versus Interleaved block type conditions was counterbalanced across two days of testing, such that half of participants underwent the Blocked condition on Day 1 and the Interleaved condition on Day 2, and the other half underwent the Interleaved condition first. Trials in both the Interleaved and Blocked conditions were presented across nine blocks of trials per day with 84 trials in each block (12 trials per coherence level in each block). In the Blocked condition, dot density was pseudorandomly assigned to block number, subject to the constraints

that (1) blocks 1-3, 4-6, and 7-9 contained one each of the Low, Medium, and High dot density conditions, and (2) density could not be identical across consecutive blocks.

Overall, participants completed 756 trials total in each of the Blocked and Interleaved conditions, with 36 trials for each combination of trial-order (Blocked / Interleaved), dot density (Low / Medium / High), and motion coherence (7 levels in total, spaced evenly between 10% and 80% coherence). Each day of testing lasted about an hour and 15 minutes, such that participants underwent about 2.5 hours of testing in total. Day 2 occurred between 1 - 3 days after Day 1. A schematic of block structure is shown in Figure 2B.

Prior to testing on each day, participants performed at least one block of practice trials (and possibly more depending on the discretion of the experimenter, who monitored participant performance during practice to ensure adequate understanding and performance of the task). During practice, participants engaged in the same task as the main task, but also received trial-by-trial auditory feedback regarding the correctness of their responses (high tone for correct, low tone for incorrect). Practice blocks contained 12 trials in which the three levels of dot density were pseudorandomly interleaved (even on Blocked condition days), with motion coherence set to 100%. During the entirety of the first 6 trials of a practice block, red circles were shown around the edges of the left and right regions of coherence in order to familiarize the participant with what regions of the screen could potentially contain coherent motion. The practice was designed to allow participants to become comfortable with the task and response options, and to ensure they understood the task and key mappings for choices and confidence ratings.

All behavioral procedures were programmed in PsychToolbox and implemented on a MacBook Pro with OSX Version 10.9.5 running Matlab r2013b.

## 2.5. Data analysis

### *2.5.1. Fitting the type 2 psychometric curve for confidence vs d'*

To assess how the relationship between confidence and d' was modulated by motion coherence, dot density, and block type, we modeled the relationship between confidence and d' with the logistic function with location parameter μ and scale parameter s, scaled and translated so as to have a range on the interval [1, 4]:

$$conf = f(d' \mid \mu,\ s) \ = 3\ \left(\frac{1}{1+e^{-(d'-\mu)/s}}\right) + 1 \qquad (1)$$

We fit the logistic function relating confidence and d' separately for each level of dot density and block type for each subject, and then submitted the parameters μ and s to 3 (dot density) x 2 (block type) repeated measure ANOVAs. The parameter s is a scaling factor determining the slope of the confidence vs d' curve. The parameter μ is of particular interest, as it corresponds to the d' value at which confidence = 2.5, the midpoint of the 4-point rating scale, and therefore

is a measure of the threshold of the confidence vs d' curve. Curves with higher confidence across d' levels have a smaller d' value at which confidence = 2.5, and thus a smaller value of μ. Thus, we expected to find a dot density (Low, Medium, High) x block type (Blocked vs Interleaved) interaction on μ, such that the effect of dot density on μ was stronger for the Interleaved condition than for the Blocked condition.

Because both confidence and d' are random variables measured with error, the metaperceptual curve fitting procedure needs to take into account the fit of the curve to both variables. We therefore measured the error of the metaperceptual curve fit as follows. For a given data pair ($d'_i$, $conf_i$), we measured error in the confidence fit as the discrepancy between $conf_i$ and $f(d'_i \mid \mu, s)$, the confidence value predicted by the curve fit at $d'_i$. More formally,

$$\varepsilon_{conf_i} = f(d'_i \mid \mu, s) - conf_i \quad (2)$$

Similarly, we measured error in the d' fit as the discrepancy between $d'_i$ and $f^{-1}(conf_i \mid \mu, s)$, the d' value predicted by the curve fit at $conf_i$:

$$\varepsilon_{d'_i} = f^{-1}(conf_i \mid \mu, s) - d'_i \quad (3)$$

where

$$d' = f^{-1}(conf \mid \mu, s) = -s \log\left(\frac{4-conf}{conf-1}\right) + \mu \quad (4)$$

Note that the logarithmic term in the above equation becomes infinite when conf = 1 or conf = 4. To avoid this, when performing curve fitting we substituted all values of conf = 1 with conf = 1.01, and all values of conf = 4 with conf = 3.99. Since each dot density x block type condition had 36 trials, the lowest and highest possible mean confidence values in a condition aside from 1 and 4 were (35*1 + 2) / 36 = 1.03 and (35*4 + 3) / 36 = 3.97. Thus, the adjusted confidence value used (3.99) when true confidence was at ceiling (4) was larger than the next highest possible true confidence value (3.97), and similarly for the adjustment when true confidence was at floor.

We computed overall error for a given ($d'_i$, $conf_i$) pair as the absolute value of the product of the errors for d' and confidence:

$$\varepsilon_i = |\varepsilon_{d'_i}\ \varepsilon_{conf_i}| \quad (5)$$

By taking the product of the errors in d' and confidence, we sidestepped the issue that d' and confidence are measured in different units and range over different scales, since residuals measured in different units can be combined if multiplication is used instead of addition to define the error term (Samuelson, 1942).

Metaperceptual curve fitting thus proceeded by finding the values of μ and s that minimized the sum of errors across all motion coherence levels, $\sum_i \varepsilon_i$. Fitting was performed using the fmincon function of Matlab.

We further used the results of the metaperceptual curve fitting to compute what values of confidence would be predicted for d' values of [0.5, 1, 1.5, …, 3], for every dot density x block type condition of every subject. In this case, the value of d' is known exactly and set to the same value for every subject, and so can be treated as an independent variable. We then submitted these predicted confidence values to a 6 (d') x 3 (dot density) x 2 (block type) repeated measure ANOVA. The purpose of this analysis was to give additional insight on how the dependence of confidence on dot density and block type might be modulated by d'.

*2.5.2. Model-free statistical approach*

To quantify the difference in confidence as a function of matched performance (d') values across pairs of dot density levels, we turn next to a model-free statistical approach as complement to and confirmation of findings from the metaperceptual curve approach developed above. For each of the Medium and High density levels, we calculated the difference in confidence rating between that level and the Low density level as a function of the difference in performance (d') for all possible pairs of performances. That is, for dot density $D \in \{High, Medium\}$, d' value $i$ at that density level for that coherence (7 levels of d' total, one at each coherence level), and d' value $j$ at each coherence level within the Low density level, we calculated

$$\delta_{D,i,j} = d'_{D,i,j} - d'_{Low,i,j} \quad (6)$$

and

$$C_{D,i,j} = confidence_{D,i,j} - confidence_{Low,i,j} \quad (7)$$

We calculated $C_D$ and $\delta_D$ separately for blocked versus interleaved trials. Finally, by plotting $C_D$ against $\delta_D$ for each subject, we obtained the difference in confidence elicited by difference in dot density (High-Low, Medium-Low) defined as the y-intercept of a fitted linear regression line ($C_D = \beta_0 + \beta_1 \delta_D$), i.e., the confidence difference $C_D$ at matched performance ($\delta_D = 0$). We then examined whether these y-intercepts significantly differed from each other and from 0 using a 2 (block type) x 2 repeated-measures ANOVA (High-Low vs Medium-Low) followed by one-sample t-tests of each set of y-intercepts ($\beta_0$).

# 3. Results

3.1. Examining psychometric curve fits for confidence vs d'

The first set of analyses examined the results of fitting type 2 psychometric curves to the confidence and d' data as described in the Methods. Curve fitting was conducted using the logistic function (Eq. 1) and thus yielded two fitted parameters of interest, μ and s.

We first examined μ, the location parameter of the logistic function. In the context of our curve fitting procedure, μ corresponds to the value of d' at which confidence achieves its mid-point value of 2.5 on the 4-point rating scale, and thus serves as a measure of the threshold of the metaperceptual function relating confidence to d'. Curves with higher confidence across d' levels have a smaller d' value at which the mean confidence rating equals 2.5, and thus a smaller value of μ. Thus, if the factors of dot density and block type affect performance-matched confidence across a wide range of d' values, they should exhibit statistically significant effects on μ.

A 3 (dot density) x 2 (block type) repeated measures ANOVA on μ revealed no main effect of block type ($F(1,20) = 0.004$, $p = 0.95$), showing that the d' required to reach the mid-point confidence threshold is similar between Interleaved versus Blocked conditions. However, there was a main effect of dot density ($F(2,40) = 15.23$, $p = 1e\text{-}5$), corresponding to the fact that higher density conditions require lower d' in order to achieve the mid-point confidence threshold (or, equivalently, that for a fixed level of d', higher dot density leads to higher confidence). We also observed a block type x dot density interaction ($F(2,40) = 3.76$, $p = 0.032$), showing that the strength of the dot density effect (i.e., higher confidence for higher dot density) differed for the Interleaved and Blocked conditions. To explore this interaction, we performed follow-up ANOVAs separately for each block type. For the Interleaved condition, we observed a main effect of dot density ($F(2,40) = 15.08$, $p = 1e\text{-}5$), again corresponding to a smaller mid-point confidence threshold under higher density. However, this main effect of dot density was not significant in the Blocked condition ($F(2,40) = 1.18$, $p = 0.32$), suggesting that the effect of dot density on confidence is considerably weaker when density is blocked rather than interleaved.

We next examined s, the scaling parameter of the logistic function which controls the slope of the metaperceptual function. A 3 (dot density) x 2 (block type) repeated-measures ANOVA on s revealed a marginal main effect of block type ($F(1,20) = 3.11$, $p = 0.09$), showing that slope is marginally higher in the Blocked condition. We also observed a significant main effect of dot density ($F(2,40) = 5.16$, $p = 0.01$), such that slope was significantly higher when dot density was higher. We observed no significant interaction between block type and dot density ($F(2,40) = 0.98$, $p = 0.4$). Together, these findings suggest that higher dot density leads to a bigger increase in confidence for each unit increase in perceptual performance (d'), but that this effect is similar across Interleaved and Blocked conditions. Average values for confidence and d' as a

function of block type, dot density, and motion coherence, and the corresponding logistic fits to the metaperceptual curves, are visualized in Figure 3.

We conducted an alternative version of this curve-fitting analysis in which, in addition to the logistic parameters μ and s, a third free parameter $c_b$ was introduced in order to allow confidence in the curve fits to span over the range [$c_b$, 4] rather than the fixed range [1, 4]. The rationale for this alternative analysis was that it might provide better fits to the data, given that participants tended to have mean confidence > 1 even when d' = 0. This approach yielded qualitatively similar fits and statistical results, but also led to some implausible patterns in single-subject fits. For full details, see Supplementary Material Section S2 and Figures S2 - S5.

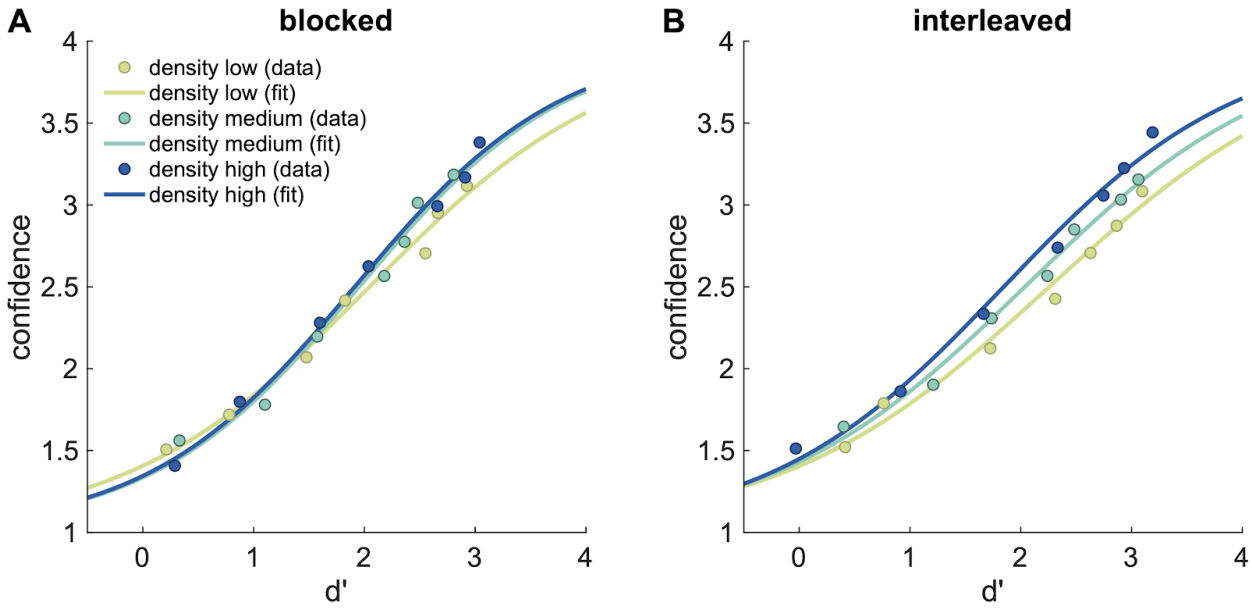


**Figure 3.** Metaperceptual curves relating confidence to task performance show how performance-matched confidence is modulated by dot density and trial order. Participants' perceptual task performance (d') increased monotonically with motion coherence; here, we set aside motion coherence and treat d' as a predictor variable for confidence (Figure 1). For fixed levels of d', confidence increased monotonically with dot density across a wide range of d' values (main effect of dot density on confidence threshold μ, p = 1e-5), particularly for higher d' values (see also Figure 4). This effect was more pronounced when dot density levels changed randomly across trials ("interleaved," right panel) rather than being held constant within a block ("blocked," left panel) (dot density x trial-order interaction, p = 0.032). Statistics were performed on single-subject curve fits, but for purposes of illustration the plots here show curve fits to the group-averaged confidence vs d' data. See Figure S1 for plots of the averages of metaperceptual curves fitted at the single-subject level, which are qualitatively similar.

### 3.2. Interlude

In the above analysis, the logistic function parameters μ and s capture the relationship between confidence and d', which are both random variables measured with uncertainty. It would also be of interest to treat d' as an independent variable and investigate the joint effects of d', dot density, and block type on confidence. However, it is not feasible to take this approach in a straightforward way, since (1) d' is a dependent variable measured with error, and (2) d' values are not perfectly matched across levels of dot density and motion coherence, which would therefore produce confounds in the analysis of confidence since confidence depends on d'. In the remaining sets of analyses (Sections 3.3, 3.4, and 3.5), we adopt several different approaches that attempt to circumvent these issues to allow for more direct analysis of the effects of d', dot density, and block type on confidence. Together, these analyses complement the metaperceptual curve-fitting analysis above and round out our understanding of how confidence depends on d', dot density, and block type.

### 3.3. Using metaperceptual curve fits to investigate the joint effects of d', dot density, and block type on confidence

One alternative way to investigate the data is to use the single-subject metaperceptual curve fits to produce predictions for what confidence would be at fixed d' values, and then analyze these predicted confidence values as a function of d', dot density, and block type. Importantly, since the input d' is known exactly and can be fixed across subjects, this method allows for treating d' as an independent variable. Thus, for every subject, we used the metaperceptual curve fits to produce predicted confidence for d' values of [0.5, 1, 1.5, ... , 3] and submitted these predicted confidence values to a 6 (d') x 3 (dot density) x 2 (block type) repeated measure ANOVA. This analysis revealed similar if more nuanced results to the above analysis. First, we observed no main effect of block type ($F(1,20) = 0.05$, $p = 0.8$), indicating that overall Blocked and Interleaved conditions produced similar confidence levels. However, we did observe a main effect of dot density ($F(2,40) = 13.85$, $p = 3e\text{-}5$), showing that higher dot density led to higher confidence judgments overall. We also observed a modest block type x dot density interaction ($F(2,40) = 3.15$, $p = 0.054$), corresponding to the stronger effect of density on confidence in the Interleaved condition.

To further characterize the block type x dot density interaction, we conducted two 6 (d') x 3 (dot density) repeated measures ANOVAs on the predicted confidence derived from metaperceptual curve fits separately for each of the Interleaved and Blocked conditions. In the Interleaved condition, there was a significant main effect of dot density on confidence ($F(2,40) = 16.38$, $p = 6e\text{-}6$), which was modulated by a d' x dot density interaction ($F(10,200) = 5.94$, $p = 7e\text{-}8$). By contrast, in the Blocked condition the main effect of dot density was not significant ($F(2,40) = 0.56$, $p = 0.6$), although the effect of density was modulated by a d' x dot density interaction ($F(10,200) = 2.20$, $p = 0.019$)

This analysis approach allows us to extend beyond the observations of the primary metaperceptual curve-fitting analyses by examining main effects of d' and interactions with this factor. The first observation is a significant main effect of d' ($F(5,100) = 290.24$, $p = 7e\text{-}58$), which shows the expected result that confidence increases with perceptual performance capacity. We also observed a d' x block type interaction ($F(5,100) = 2.96$, $p = .016$) and a d' x dot density interaction ($F(10,200) = 10.08$, $p = 1e\text{-}13$); this mirrors the main effects of block type and dot density in the earlier ANOVA on the fitted logistic function parameter s, showing that for a given d' level, confidence is higher in the Blocked than the Interleaved condition, and the effect of dot density on confidence grows for larger d' values (Figure 4). This observation has interesting implications for the design of experiments probing the MPDC effect; we discuss this point in greater detail in the Discussion. Finally, we note that the absence of a d' x block type x dot density interaction ($F(10,200) = 0.52$, $p = 0.9$) suggests that the growing dependence of confidence on dot density as d' increases is not strongly different for the Blocked versus Interleaved conditions: both block types show the metaperceptual curves for each level of dot density separating more in confidence as d' increases (Figure 4).

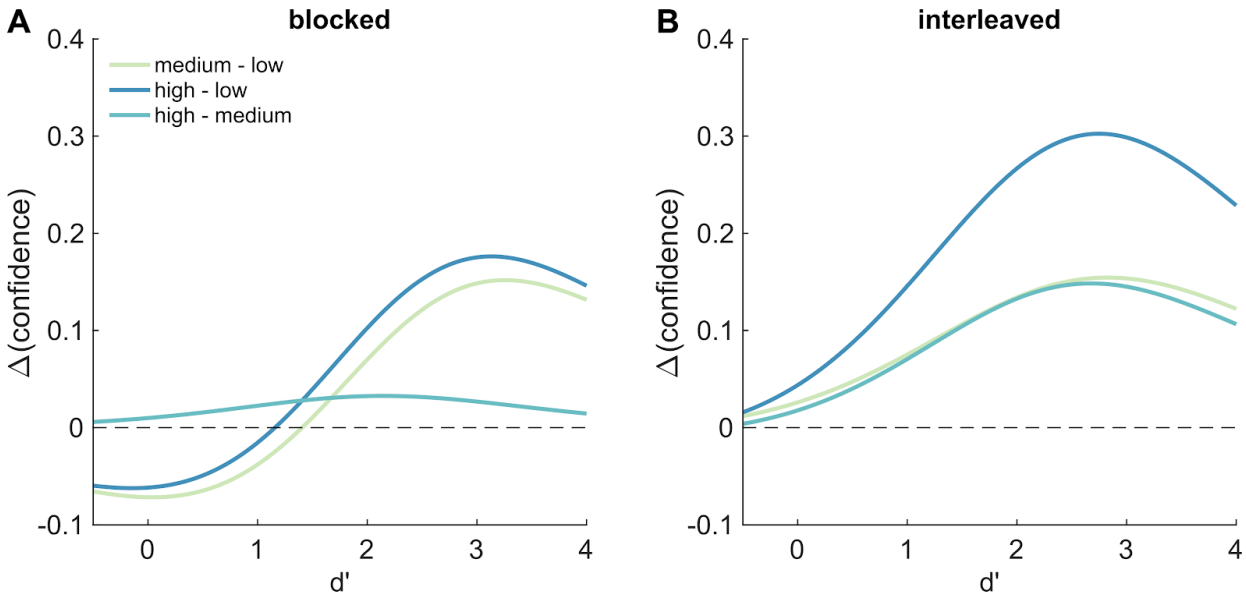


**Figure 4.** Difference curves for metaperceptual curves fitted to group-average data. These plots show difference curves for the metaperceptual curve fits to the group-averaged confidence vs d' data, as depicted in Figure 3. The difference curves highlight how the effect of dot density on confidence changes across levels of d', being negligible when d' = 0 and steadily growing as d' increases. This suggests that experiments seeking to investigate MPDC effects should probe relatively high values of d' of 2.5 (corresponding to ~90% correct responding for an unbiased observer) or higher in order to maximize the magnitude of MPDC effects. Note that since average d' in our experiment maxed out at about 3 (Figure 3), it is unclear if the decrease in confidence differentials in the d' > 3 range reflects a true effect or an artifact of curve fitting. These difference curves also suggest that in the Interleaved condition, multiplicative increases in dot density (which increased by a factor of 3 across each level of density) yielded additive increases in performance-matched confidence.

## 3.4. Investigating MPDC effects in conditions that exhibit approximately matched d'

### *3.4.1. Identifying performance-matched coherence conditions*

Another way to circumvent the analysis issues mentioned in the Interlude (Section 3.2) is to *post-hoc* select the experimental conditions that happened to yield roughly equivalent levels of d', and then investigate the effects of motion coherence, dot density, and block type on confidence in these conditions. Thus, for this approach, we first conducted analyses to identify conditions in which d' was approximately matched.

A 7 (motion coherence) x 3 (dot density) x 2 (block type) ANOVA on d' revealed a main effect of coherence ($F(6,120) = 267.15$, $p = 6e\text{-}67$) but no main effects of block type ($F(1,20) = 0.77$, $p = 0.4$) or dot density ($F(2,40) = 0.51$, $p = 0.6$). We observed no interactions between block type and dot density ($F(2,40) = 1.45$, $p = 0.2$) or block type and coherence ($F(6,120) = 1.1$, $p = 0.4$), nor a 3-way interaction between block type, dot density, and coherence ($F(12,240) = 1.0$, $p = 0.4$).

However, we did observe a significant coherence x dot density interaction ($F(12,240) = 2.97$, $p = 0.0007$), indicating that not all coherence levels exhibited perfect performance (d') matching across dot densities. To explore this interaction, we performed several one-way ANOVAs on d' with dot density as a factor, separately for each block type and coherence level, to seek block type and coherence conditions where d' is "well enough" matched (i.e., conditions in which there was no significant effect of dot density on d'). Note that since we decline to correct for multiple comparisons here, our approach is conservative in identifying block type and coherence conditions where d' is well-matched.

The p-values for the effect of dot density on d' for each block type (2) x coherence (7) level ANOVA performed in this analysis are presented in Table 1. In the table we mark three motion coherence levels where d' appears to be roughly matched in both block type conditions.

| | **Coherence level** | | | | | | |
|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3*† | *4* | *5*† | *6* | *7*† |
| **Blocked** | 0.732 | 0.051 | 0.683 | 0.099 | 0.250 | 0.012 | 0.279 |
| **Interleaved** | 0.007 | 0.023 | 0.881 | 0.867 | 0.303 | 0.942 | 0.661 |

**Table 1.** P-values for step-down one-way ANOVAs on d' within each block type condition and coherence level. †s indicate motion coherence levels for which d' is roughly matched in both block type conditions.

We focused on these three motion coherence levels — levels 3, 5, and 7, corresponding to 33.33%, 56.67%, and 80% coherence — for the subsequent confidence analyses described below. d' values for these conditions are presented in Table 2.

| | Coherence level, Blocked | | | Coherence level, Interleaved | | |
|---|---|---|---|---|---|---|
| | **3** | **5** | **7** | **3** | **5** | **7** |
| **Low** | 1.48 | 2.55 | 2.93 | 1.73 | 2.63 | 3.10 |
| **Medium** | 1.58 | 2.36 | 2.81 | 1.74 | 2.48 | 3.06 |
| **High** | 1.60 | 2.66 | 3.04 | 1.66 | 2.75 | 3.19 |
| **Mean** | **1.55** | **2.52** | **2.92** | **1.71** | **2.62** | **3.12** |

**Table 2.** Performance (d') for the three matched-performance coherence across dot density levels.

*3.4.2. Confidence within performance-matched coherence conditions*

The mean confidence reported for each of the three motion coherence levels where performance was matched are shown in Table 3. Interestingly, multiplicative increments in dot density (recall that Low = 1 dot/deg$^2$, Medium = 3 dots/deg$^2$, High = 9 dots/deg$^2$) led to approximately additive changes in performance-matched confidence (see also Figure 4). The increases in confidence due to increased dot density were nevertheless of comparable or larger effect size to previous reports in the literature on the same 1-4 scale (confidence differences ranging approximately 0.1-0.3; compare to ~0.1-0.3 as reported by Koizumi and colleagues (2015), or ~0.1 as reported by Samaha and colleagues (2016)).

| | Coherence level, Blocked | | | Coherence level, Interleaved | | |
|---|---|---|---|---|---|---|
| | **3** | **5** | **7** | **3** | **5** | **7** |
| **Low** | 2.07 | 2.71 | 3.12 | 2.12 | 2.71 | 3.08 |
| **Medium** | 2.20 | 2.78 | 3.19 | 2.31 | 2.85 | 3.16 |
| **High** | 2.28 | 2.99 | 3.38 | 2.33 | 3.06 | 3.44 |
| **Mean** | **2.18** | **2.82** | **3.23** | **2.26** | **2.87** | **3.23** |

**Table 3.** Confidence for the three matched-performance coherence across dot density levels.

We next performed similar one-way ANOVAs within each block type (2) and coherence level (7) but this time with confidence as the outcome variable. Of interest are the p-values for the effect of dot density on confidence for each, shown in Table 4 (all d'-matched coherence levels are marked with †s as in Table 1). Cells marked with ‡s show the d'-matched coherence levels

which exhibited a significant effect of dot density on confidence in these one-way ANOVAs, relying on a Bonferroni-corrected alpha level of 0.05 / 6 = 0.008.

| | **Coherence level** | | | | | | |
|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3*† | *4* | *5*† | *6* | *7*† |
| **Blocked** | 0.006 | 0.370 | 0.006‡ | 0.018 | < 0.001‡ | 0.012 | < 0.001‡ |
| **Interleaved** | 0.009 | 0.157 | 0.017 | 0.002 | 0.002‡ | < 0.001 | < 0.001‡ |

**Table 4.** Significant p-values for for step-down one-way ANOVAs on confidence. As before, †s indicate motion coherence levels for which d' is roughly matched in both block type conditions. ‡s highlight coherence levels for which a significant effect of dot density was observed (higher density → higher confidence) despite matched performance (d').

Thus, although there is evidence that the effect of dot density on performance-matched confidence is stronger in the Interleaved than in the Blocked condition, both conditions exhibited significant effects of dot density on confidence in motion coherence levels that are matched for d'.

Also of note is that for the Interleaved condition, all coherence levels from 3 - 7 have non-significant effects of dot density on d' (all ps > 0.3) (Table 1) but significant effects on confidence (all ps < 0.02) (Table 4). If these results are robust to replication, this implies that the experimental design used here can be used to yield MPDC effects at fixed coherence levels without having to calibrate stimulus parameters in order to match d'. This outcome could allow for a significant simplification in the design of future MPDC experiments, in which it can often be difficult to appropriately calibrate stimuli so as to achieve robust MPDC effects.

*3.4.3. Analysis of confidence at matched d'*

The three motion coherence levels with matched d' can also be used to perform a separate test of the effect of coherence, dot density, and block type on d'-matched confidence.

First, we should confirm that d' is roughly matched not only across dot density but also block type by conducting 2 (block type) x 3 (dot density) ANOVAs on d' within each motion coherence level. None of the coherence levels exhibit an effect of dot density, block type, or dot density x block type on d' (ps > 0.2), so at these coherence levels d' is roughly matched not only across dot density but also block type. (As Table 2 shows, there is a slightly larger average d' in the Interleaved than in the Blocked condition, with average magnitude about 0.15).

In contrast, a 2 (block type) x 3 (dot density) x 3 (coherence) ANOVA with confidence as the outcome variable revealed the expected main effect of coherence ($F(2,40) = 218.45$, $p = 3e\text{-}22$;

higher coherence → higher confidence) and dot density (F(2,40) = 17.13, p = 4e-6; (higher density → higher confidence), but not for block type (F(1,20) = 0.34, p = 0.6). Somewhat surprisingly, however, we observed no trending interaction between block type x dot density (F(2,40) = 0.66, p = 0.5). It is possible that this null finding is partially attributable to loss of information due to omitting 4 of the 7 motion coherence levels from analysis, since statistical investigations of the block type x dot density interaction in previous analyses that incorporated all motion coherence levels (Sections 3.1 and 3.3) revealed only modestly significant effects (ps = 0.032 and 0.054, respectively). It is also worth noting that although MPDC effects appear stronger in the Interleaved condition, they may not be entirely absent from the Blocked condition for higher values of d' (see Figure 4), despite the fact that the effect of density on performance-matched confidence in the Blocked condition does not reach statistical significance in the analyses of Sections 3.1 and 3.3 (ps = 0.3 and 0.6, respectively). In this analysis we also observed a significant dot density x coherence interaction (F(4,80) = 2.47, p = 0.05), showing that MPDC effects get stronger at higher d' levels, consistent with previous analyses on the slope of the psychometric curve (Section 3.1) and the d' x dot density interaction on confidence (Section 3.3). No significant interaction between block type x coherence (F(2,40) = 1.29, p > 0.28), or block type x dot density x coherence (F(4,80) = 0.62, p > 0.5) was observed.

Further comparisons between the analyses of Sections 3.1 and 3.4 can be found in Supplementary Material (Section S3 and Figure S6).

### 3.5. Model-free statistical approach

As a final approach to understanding the data, we computed the difference in confidence expected for a given difference in performance for all participants between the high versus low (High-Low) and medium versus low (Medium-Low) dot density conditions separately for each subject (see Methods). Across all participants, these differences appeared well described by a linear relationship Figure 5A, 5B), justifying the use of fitted y-intercepts (i.e., $C_D$ where $\delta_D = 0$) to describe the expected change in confidence resulting from dot density manipulations. However, although the relationship was roughly linear across all comparisons (High-Low and Medium-Low in both Blocked and Interleaved conditions), the shift in the distribution is the critical element to evaluating the presence and magnitude of MPDC effects.

To examine this effect, we performed a 2 (Interleaved vs Blocked) x 2 (High-Low vs Medium-Low) repeated measures ANOVA on the y-intercepts of these fitted regression lines. This analysis first revealed a main effect of block type (Interleaved vs Blocked; F(1,20) = 4.62, p = .044), showing that the y-intercepts were higher in the Interleaved than the Blocked condition (Table 4) across both High-Low and Medium-Low comparisons. We also observed a main effect of density pair (High-Low vs Medium-Low; F(1, 20) = 17.63, p < .001), with High-Low having a larger y-intercept than Medium-Low across both Blocked and Interleaved conditions. Finally, we observed an interaction between block type and density pair (F(1,20) = 4.78, p = .041), meaning that the increase in y-intercept was much stronger for the Interleaved than the Blocked condition (Figure 5C).

One-sample t-tests against 0 revealed that y-intercepts differed significantly from 0 in the High-Low condition but not the Medium-Low condition (Table 5), showing that higher dot density led to higher confidence despite matched performance in both the Blocked and Interleaved conditions as long as the density difference was large enough, i.e., an MPDC effect. The magnitude of this effect (~0.05-0.25, Table 5) was also on par with the findings reported above and in previous reports in the literature on the same 1-4 scale (Koizumi et al., 2015; Samaha et al., 2016).

| | **Medium-Low** | | | | **High-Low** | | | |
|---|---|---|---|---|---|---|---|---|
| | *μ (sd)* | *Cohen's d* | *t(20)* | *p* | *μ (sd)* | *Cohen's d* | *t* | *p* |
| **Blocked** | .046 (.178) | .2574 | 1.180 | .252 | .081 (.218) | .3723 | 2.742 | .013 |
| **Interleaved** | .105 (.176) | .5984 | 1.706 | .104 | .253 (.242) | 1.0426 | 4.778 | < .001 |

**Table 5.** Means, standard deviations, effect size (Cohen's d), and results of one-sample t-tests against 0 for the fitted y-intercepts to predicted confidence differences as a function of performance differences.

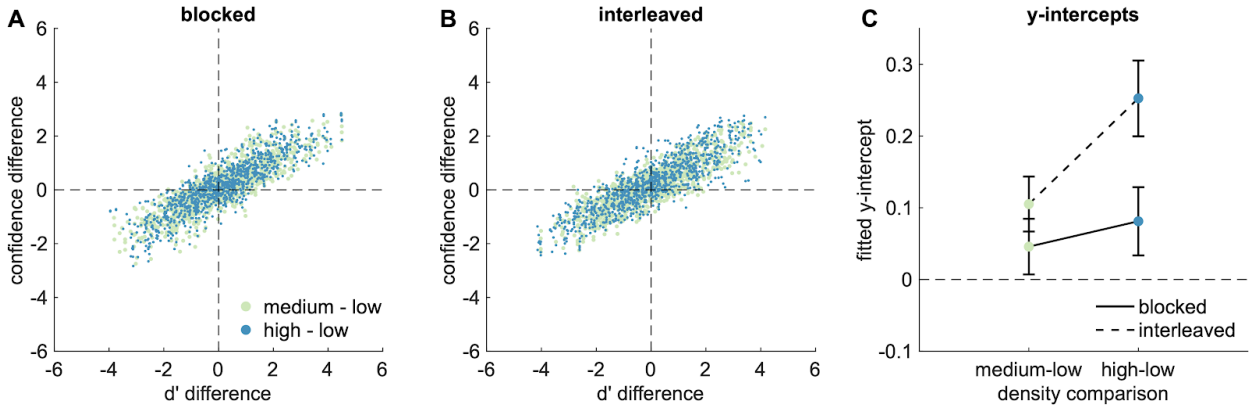


**Figure 5.** Differences in confidence as a function of difference in d' between the High-Low dot density comparisons and the Medium-Low dot density comparisons. (A) Scatterplot of difference in confidence as a function of d' difference for the Blocked condition, for both density comparison types, for all subjects. Each dot represents a single confidence comparison versus d' comparison point for a single subject. (B) Scatterplot of the same as (A) but for the Interleaved condition. (C) Fitted y-intercepts across all participants to the Blocked and Interleaved conditions, demonstrating the expected confidence difference at performance-matched conditions, i.e., $C_D$ when $\delta_D = 0$. The High-Low comparison demonstrates expected confidence differences significantly above 0 for both the Blocked and Interleaved conditions ($t_{blocked,high\text{-}low}(20) = 2.742$ and $t_{interleaved,high\text{-}low}(20) = 4.778$, $p < .001$). See main text for details.

Finally, as a control analysis, we created another model by shuffling the labels for dot density and repeating the above analysis 100 times, focusing on the results of the t-tests as these are

the critical component to demonstrate MPDC effects. We found that in all four cases (Blocked/Interleaved x High-Low/Medium-Low), the distributions of y-intercepts fitted to the shuffled data were centered almost exactly at 0 (Table 6), demonstrating that the effect is not dependent on irrelevant factors within the distribution of data, but is specific to the dot density manipulations performed here.

| | **Medium-Low** | | | **High-Low** | | |
|---|---|---|---|---|---|---|
| | *μ (sd)* | *t(20)* | *p* | *μ (sd)* | *t* | *p* |
| **Blocked** | -.038 (.236) | .741 | .467 | -.085 (.258) | 1.13 | .272 |
| **Interleaved** | -.031 (.125) | 1.517 | .145 | .015 (.219) | 0.310 | .760 |

**Table 6.** Means, standard deviations, and results of one-sample t-tests against 0 for the null (density-shuffled) model.

3.6. Qualitative replication in an online data set

Prior to collecting the laboratory data described above, we conducted pilot experiments on Amazon Mechanical Turk using a similar experimental design in order to achieve a preliminary assessment of the potential effects of motion coherence, dot density, and block type on d' and confidence. Results from these experiments are summarized in Supplementary Material (Section S5) and depicted in Figure S8. In spite of noisy data and low sample size, the results were qualitatively similar to the ones described above, and thus constitute a modest qualitative replication of the main findings: across a broad range of d' values, performance-matched confidence increases as a function of dot density, and this effect is stronger when density is interleaved rather than blocked. See Section S5 for full discussion.

## 4. Discussion

Here we present, for the first time, a full psychometric characterization of the relationship between task performance and confidence in data exhibiting matched-performance / different-confidence (MPDC) effects, which are crucial tools for studying confidence and awareness independently of performance confounds (Morales et al., 2019). We call the curve relating meta-awareness to perception the type 2 psychometric function (by way of analogy to the conventional psychometric function relating perception to stimulus properties), and also introduce the related term "metaperceptual" as a shorthand way to refer to the relationship between meta-awareness and perception (by way of analogy to the term "psychophysical") (Figure 1; see figure legend for further discussion). By measuring the full type 2 psychometric function, we were able to address several questions of practical importance for the understanding and application of MPDC effects, including: How does the magnitude of MPDC effects depend on the level of d' at which performance is matched? How does the magnitude of MPDC effects depend on the strength of the psychophysical manipulation used to achieve them

(here, dot density)? How does the effectiveness of MPDC-inducing psychophysical manipulations depend on their temporal organization (here, probed by comparing Blocked and Interleaved ordering of dot density)? And, how can we best quantify these effects using a principled analytic approach?

In our task, participants viewed full-screen displays of moving dots and had to judge whether coherent downward motion emerged in a region on the left or right side of the screen, and then rate confidence in the accuracy of their perceptual decision. Motion coherence varied across trials over a wide range spanning 10% to 80% coherence, which allowed us to construct full type 2 psychometric functions spanning performance levels ranging from near-chance levels (d' $\cong$ 0) to near-ceiling levels (d' $\cong$ 3.5, which corresponds to 96% correct responding for an unbiased observer). Following previous demonstrations (Koizumi et al., 2015; Odegaard et al., 2018; Rollwage et al., 2020), we varied dot density across trials to independently manipulate confidence, and assessed how the effect of this manipulation on performance-matched confidence varied as a function of task performance and temporal organization (by comparing Blocked vs Interleaved groupings of dot density).

Our approach to analyzing the data exemplifies several complementary methods for analyzing MPDC effects with type 2 psychometric functions. First, we conducted metaperceptual curve fitting (Section 3.1) using a modified form of the logistic function (Eq. 1), being careful to measure error in a scale-invariant manner (Eq. 5) since the metaperceptual curve must be fit to both the x-axis (d') and y-axis (confidence) variables. We then analyzed the fitted parameter values to make inferences about how our experimental manipulations influenced the overall behavior of the metaperceptual function. Using the curve fits, we could also generate predicted confidence values for given d' values, allowing for a complementary analysis of performance-matched confidence across the entire metaperceptual function (Section 3.3).

We complemented these parametric analysis approaches with two non-parametric approaches. First, we investigated confidence effects in motion coherence levels that happened to exhibit roughly matched levels of d' (Section 3.4), mirroring studies that seek to directly match performance across conditions. We furthermore performed a model-free regression analysis (Section 3.5) following precedent set by Knotts and colleagues (2018), examining differences in confidence elicited by dot density manipulations as a function of differences in performance.

Altogether, these complementary approaches to analyzing MPDC effects in the type 2 psychometric function allowed us to advance our understanding of the behavior of MPDC effects in several ways, as summarized below. These insights can help inform the design of future experiments seeking to use MPDC effects to study the nature of meta-awareness.

4.1. MPDC effects are stronger when MPDC-inducing stimulus manipulations are interleaved rather than blocked

Perhaps the most striking finding of this study was the strong dependence of MPDC effects on the temporal organization of the stimulus manipulation used to induce it. Analysis of type 2 psychometric curve fits revealed that changes in dot density were more effective at driving changes in confidence when dot density changed frequently and unpredictably from trial to trial (Interleaved condition), as compared to situations where dot density was constant and predictable throughout a block of trials (Blocked condition) (Figures 3, 4). A complementary, model-free analysis combining data across the entire metaperceptual function yielded similar conclusions (Figure 5), and we observed similar findings in a pilot study run via the online platform Amazon Mechanical Turk (see Supplementary Material Section S5 for details).

One possible explanation for these findings hinges on the dynamics of type 2 criterion setting, i.e., the rules and standards by which an observer decides how to produce a confidence rating given a certain level of perceptual evidence. Decision criteria are malleable and can change according to context, as revealed by decades of research on signal detection theory (Macmillan & Creelman, 2004). For instance, suppose an observer must make a perceptual decision based on a moderate level of perceptual evidence $e_i$ and then rate confidence in that decision. The observer might be more likely to endorse this decision with high confidence (e.g., a rating of 3 or 4 on a 4-point scale) if it was made in the context of recent trials being much more difficult, i.e., if $e_j \ll e_i$ for most j. By contrast, if the same decision based on the same perceptual evidence was made in the context of recent trials being much easier ($e_j \gg e_i$ for most j), the observer might be more likely to report lower confidence (e.g., 1 or 2 on a 4-point scale).

In the same way, even if high dot density tends to induce high confidence, this effect could be relatively masked or “washed out” when high density trials are grouped together in blocks. Since *all* trials in the block would have high dot density (and thus relatively higher confidence), the *typical* level of confidence experienced in the block would be higher, and so render high confidence on any particular trial as less remarkable in context and thus less likely to receive the highest confidence ratings.

A related observation is that human participants have difficulty dynamically adjusting response criteria from trial to trial in response to frequent changes to stimulus characteristics, even when it would be optimal to do so (Brown and Steyvers 2005; Gorea and Sagi 2000; Adler and Ma 2018). Thus, randomly interleaving dot density across trials likely helps to ensure that participants use a fixed decision strategy for rating confidence across dot density levels, thus placing confidence ratings across density levels on a more even footing. By contrast, organizing dot density by blocks of trials allows participants to more easily change decision strategy for rating confidence across blocks, thereby potentially obscuring true differences in confidence across density levels.

It is also worth noting that, although MPDC effects were stronger in the Interleaved condition, nonetheless we also observed some evidence for MPDC effects even in the Blocked condition. The model-free analysis (Section 3.5, Figure 5) revealed a weak but significant effect of dot density on confidence in the Blocked condition, and we also found confidence to be significantly modulated by density for individual motion coherence levels exhibiting roughly matched levels of d' across density (Section 3.4). However, analyses based on the full metaperceptual function failed to find a significant effect of dot density on performance-matched confidence in the Blocked condition (Sections 3.1 and 3.3), perhaps because an MPDC effect is not evident for lower d' values in the Blocked condition (Figures 3, 4, S6).

### 4.2. MPDC effects are more pronounced at higher levels of d'

We found that the magnitude of MPDC effects was dependent on perceptual task performance (d'), such that higher levels of d' were associated with a more pronounced increase in confidence as a result of increase in dot density (Figures 3, 4, S6). This entails that future experiments implementing MPDC manipulations should seek to do so at high values of d', ideally in the range of 2.5 - 3 (corresponding to 89% - 93% correct responding for an unbiased observer), in order to maximize the magnitude of confidence differences across performance-matched conditions. Naturally, this consideration should be counterbalanced by a consideration of available resources — error trials become increasingly rare as d' increases, entailing that more trials are required to reliably estimate hit rate and false alarm rate (and therefore, d'). Thus, for designs with relatively low trial counts, choosing a somewhat smaller value of d' at which to achieve performance matching could potentially be the best choice, all things considered.

Inspection of the metaperceptual function difference curves (Figure 4) seems to suggest that at very high values of d' (> 3), the magnitude of MPDC effects trails off and begins to decrease. However, the highest levels of d' achieved in the data also maxed out at about d' = 3 (see Figure S6 for a direct comparison of metaperceptual difference curves and individual data points), and so the current data set is not ideal for inferring how the metaperceptual function might behave at such high values of d'; we leave consideration of this matter to future work. However, it is worth noting that at *some* point we should expect to see metaperceptual difference curves diminish in magnitude, simply due to the fact that as d' continues to increase, confidence in all conditions should begin to approach ceiling levels (e.g. confidence = 4 on the 4-point rating scale), leaving less room for differences across conditions to manifest.

### 4.3. MPDC effects are logarithmically related to stimulus manipulations

We probed three levels of dot density in this experiment, where low, medium, and high density conditions had 1, 3, and 9 dots/deg$^2$, respectively. Thus, dot density followed a geometric progression. By contrast, in the Interleaved condition, we observed that performance-matched confidence increased in additive increments across density conditions, i.e., conf(high density) - conf(medium density) = conf(medium density) - conf(low density) (Figure 4B; Table 3). Thus,

changes in performance-matched confidence depended logarithmically upon changes in dot density, echoing various well-known logarithmic relationships between perception and stimulus properties as encapsulated by the Weber-Fechner law (Fechner et al., 1966).

By contrast, this relationship did not seem to hold in the Blocked condition. Type 2 psychometric curve fits suggested that although performance-matched confidence differences for the (medium density - low density) contrast were roughly comparable in magnitude for Blocked and Interleaved conditions, performance-matched confidence differences for the (high density - medium density) contrast were considerably smaller (near zero) in the Blocked condition (Figure 4A; Table 3). In light of the discussion in Section 4.1, this may suggest that decision strategies for rating confidence in the Blocked condition were particularly susceptible to change under high dot density, thus effectively masking across-condition changes in confidence.

Notably, this finding has consequences for minimizing stimulus confounds in MPDC experiments. Experiments that achieve MPDC effects by stimulus manipulations (such as the dot density manipulation used here) eliminate performance confounds (i.e., match d' across conditions) at the expense of creating stimulus confounds (e.g. the different stimuli used in the different dot density conditions). Naturally, it is desirable to minimize stimulus confounds to the greatest extent possible. If the logarithmic relationship between stimulus manipulation and MPDC magnitude is robust, this entails that the same MPDC effect between two conditions can be achieved with smaller corresponding stimulus confounds if the overall stimulus magnitudes utilized are smaller. For instance, in the current study, the MPDC effect for medium vs low and high vs medium dot density conditions is equivalent (Figure 4), but the stimulus confound is three times smaller in the medium vs low contrast (3 vs 1 dots/deg$^2$) than in the high vs medium contrast (9 vs 3 dots/deg$^2$). Thus, if a future study employed a similar design but with only two levels of dot density, choosing dot densities of 1 vs 3 dots/deg$^2$ would be preferable to choosing dot densities of 3 vs 9 dots/deg$^2$ due to achieving the same MPDC magnitude in spite of having a smaller stimulus confound.

### 4.4. MPDC effects at individual performance levels may occur naturally with the RDK paradigm used here

One motivation for constructing full metaperceptual functions in order to dissociate meta-awareness from task performance is that it can be difficult to achieve precise performance matching in psychophysics experiments. Typically, the experimenter attempts to ensure performance matching across conditions by suitably titrating stimulus properties prior to the main experiment. However, even well-controlled titration procedures can sometimes yield noisy results, and even small differences in task performance can significantly obscure the interpretation of MPDC experiments. Constructing a full type 2 psychometric function circumvents this concern by characterizing metaperceptual performance across a broad range of d' values. Importantly, this approach obviates the need to achieve precise performance matching for any particular pair of data points, since performance-matched confidence

differences can emerge from the metaperceptual function fit achieved by considering behavior across a broad range of d' values.

Nonetheless, it is interesting to note that in the Interleaved condition, d' did not significantly differ across dot density levels for motion coherence levels 3 - 7 (corresponding to motion coherences between 33.3% and 80%; ps > 0.3, Table 1), and yet confidence was significantly modulated by density for all these coherence levels (uncorrected ps < 0.02, Table 4) with effect size similar in magnitude to previous reports, i.e., on the order of confidence differences at matched performance levels in the range of ~0.1-0.3 (Koizumi et al., 2015; Samaha et al., 2016). Thus, it appears that the experimental design used here may be effective in achieving MPDC effects at individual motion coherence levels, even in the absence of complicated and fallible stimulus titration procedures. It may be of interest for future research to confirm whether this apparent effect is indeed robust; if so, it would provide a simple and efficient way to achieve MPDC effects at a single level of task performance.

### 4.5. A simple signal detection theory account of the effect of dot density on confidence

Why is it that higher dot density yields higher confidence for a fixed level of d'? One possible explanation is that under higher dot density, perceptual evidence for coherent motion might become noisier. For instance, suppose that on a given trial, coherent downward motion occurs in the region of coherence to the left of fixation. Relative to lower dot density conditions, higher density conditions will feature a larger number of downward moving dots in the left region of coherence. However, higher density conditions will also have a larger number of incoherent dots moving randomly in all directions, and this leads to more variability in randomly occurring downward motion in both the left and right regions of coherence (and thus noisier perceptual evidence for downward motion in both the left and right regions). In turn, perceptual evidence that is more variable is more likely to achieve higher absolute values and therefore more likely to exceed decision criteria for reporting high confidence. This link between perceptual evidence variability and confidence can be formally characterized in signal detection theory (Morales et al., 2019) and provides a natural explanatory mechanism for empirical dissociations between task performance and confidence observed across several studies (Cortese et al., 2016; Koizumi et al., 2015; Odegaard et al., 2018; Peters, Fesi, et al., 2017; Rahnev et al., 2011; Rahnev, Maniscalco, et al., 2012; Samaha et al., 2016, 2019; Stolyarova et al., 2019). In Supplementary Material Section S4 and Figure S7, we present some signal detection theory simulations based on this principle which can heuristically account for some features of our data set. Future work could employ a more formal model comparison approach to more rigorously assess candidate computational models for the empirical effect of dot density on confidence.

4.6. Applications to metacognition and consciousness research

Our focus here on using type 2 psychometric functions to investigate MPDC effects is situated within the broader context of approaches to researching metacognition and consciousness that emphasize the importance of minimizing confounds, particularly task performance confounds (Lau, 2008; Lau & Passingham, 2006; Morales et al., 2015, 2019; Peters et al., 2016; Rollwage et al., 2020; Stolyarova et al., 2019). Elevated levels of confidence, as well as the related phenomenon of conscious awareness (Lau, 2019; Rosenthal, 2019; Sherman et al., 2015; Zehetleitner & Rausch, 2013), are typically accompanied by elevated levels of task performance, but not *necessarily* so; the two can dissociate (Koizumi et al., 2015; Lau & Passingham, 2006; Li et al., 2018; Maniscalco et al., 2016; Odegaard et al., 2018; Peters, Fesi, et al., 2017; Rahnev, Bahdo, et al., 2012; Rahnev et al., 2011; Rahnev, Maniscalco, et al., 2012; Rollwage et al., 2020; Samaha et al., 2016, 2017; Solovey et al., 2015; Stolyarova et al., 2019). This entails that (1) distinct computational and neural mechanisms underlie perceptual task performance and meta-awareness (Maniscalco et al., 2019; Peters, Thesen, et al., 2017), and (2) if specific care is not taken to dissociate task performance from meta-awareness, then task performance poses a potential confound such that computational or neural mechanisms attributed to meta-awareness may in fact be better attributed to task performance. Thus, in order to make sharp inferences about meta-awareness *per se*, it must be experimentally isolated from performance confounds (Morales et al., 2015, 2019; Peters, Kentridge, et al., 2017).

Experimental MPDC effects are a primary method for achieving such experimental isolation of meta-awareness, and the work presented here aims to facilitate better understanding and usage of MPDC effects for the sake of better understanding metacognition and consciousness. Our work accomplishes this by both helping to guide the design of future experiments that seek to match performance across experimental conditions, as well as introducing the alternative approach of measuring and analyzing entire metaperceptual curves, which obviates the need to achieve precise performance matching at specific performance levels (see also Section 4.4).

Further, although the link between confidence and conscious awareness is complex (Rosenthal, 2019), we note that the range of task performance over which MPDC effects were demonstrated here may have important implications for the study of conscious versus unconscious processing at near-threshold levels of perceptual performance. In the present study we observed that the magnitude of MPDC effects is likely to shrink as d' approaches 0. It is possible that this observation may also relate to reports of confidence (or awareness) being indistinguishable from floor at low but above-chance levels of d', i.e. blindsight-like behavior (Weiskrantz, 1986). Future studies may compare the magnitude of MPDC effects near chance performance to the effect size of observed divergence between performance and confidence at low levels of d' as predicted by ideal observer models (Knotts et al., 2018; Peters, Fesi, et al., 2017; Peters & Lau, 2015; Rajananda et al., 2018).

## 4.7. Conclusions

Here we show that measuring and analyzing whole type 2 psychometric functions can enhance our understanding and application of matched-performance / different-confidence effects. In turn, these conceptual and methodological advances can be used in future research to better isolate metacognition and consciousness from performance confounds and thereby further our understanding of the underlying computational and neural mechanisms.

## Acknowledgements

This project was supported by a subaward grant from the Duke University Summer Seminars in Neuroscience and Philosophy, sourced from the John Templeton and Templeton World Charity Foundations (to BM, BO, JM, & MAKP). OGC was supported by the University of California Riverside MARC U STAR program (Award Number T34GM062756 from the National Institute of General Medical Sciences). JM was supported by the Johns Hopkins University Office of the Provost. MAKP was supported by the Canadian Institute for Advanced Research Azrieli Program in Brain, Mind & Consciousness Global Scholars Program.

## References


Azzopardi, P., & Cowey, A. (1997). Is blindsight like normal, near-threshold vision? *Proceedings of the National Academy of Sciences*, *94*(December), 14190–14194.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*(4), 412–428.

Brown, S., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(4), 587–599.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, *7*, 13669.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology*, *132*(Pt 9), 2531–2540.

Fechner, G. T., Boring, E. G., & Howes, D. H. (1966). *Elements of psychophysics*. Holt, Rinehart and Winston.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 12380–12384.

Knotts, J. D., Lau, H., & Peters, M. A. K. (2018). Continuous flash suppression and monocular pattern masking impact subjective awareness similarly. *Attention, Perception & Psychophysics*.

Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive

control? *Attention, Perception & Psychophysics*, *77*(4), 1295–1306.

Lau, H. (2008). Are We Studying Consciousness Yet? In L. Weiskrantz & M. Davies (Eds.), *Frontiers of Consciousness* (pp. 2008–2245). Oxford University Press.

Lau, H. (2019). *Consciousness, Metacognition, & Perceptual Reality Monitoring*. https://doi.org/10.31234/osf.io/ckbyf

Lau, H., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, *103*(49), 18763–18768.

Li, M. K., Lau, H., & Odegaard, B. (2018). An investigation of detection biases in the unattended periphery during simulated driving. *Attention, Perception & Psychophysics*, *80*(6), 1325–1332.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Taylor & Francis.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. K. (2019). Tuned normalization in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. In *bioRxiv* (p. 558858). https://doi.org/10.1101/558858

Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception & Psychophysics*. https://doi.org/10.3758/s13414-016-1059-x

Miyoshi, K., & Lau, H. (2020). A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychological Review*. https://doi.org/10.1037/rev0000184

Morales, J., Chiang, J., & Lau, H. (2015). Controlling for performance capacity confounds in neuroimaging studies of conscious awareness. *Neuroscience of Consciousness*, *2015*(1),

niv008.

Morales, J., Odegaard, B., & Maniscalco, B. (2019). *The Neural Substrates of Conscious Perception without Performance Confounds*. https://doi.org/10.31234/osf.io/8zhy3

Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. A. K., Lau, H., & Basso, M. A. (2018). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), E1588–E1597.

Peters, M. A. K., Fesi, J., Amendi, N., Knotts, J. D., Lau, H., & Ro, T. (2017). Transcranial magnetic stimulation to visual cortex induces suboptimal introspection. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *93*, 119–132.

Peters, M. A. K., Kentridge, R. W., Phillips, I., & Block, N. (2017). Does unconscious perception really exist? Continuing the ASSC20 debate. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix015

Peters, M. A. K., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife*, 10.7554/eLife.09651.

Peters, M. A. K., Ro, T., & Lau, H. (2016). Who's afraid of response bias? *Neuroscience of Consciousness*, *2016*(1), niw001–niw001.

Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., & Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour*.

Rahnev, D., Bahdo, L., de Lange, F. P., & Lau, H. (2012). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology*, *108*(5), 1529–1536.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F. P., & Lau, H. (2011). Attention

induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 1513–1515.

Rahnev, D., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of Neurophysiology*, *107*, 1556–1563.

Rajananda, S., Zhu, J., & Peters, M. A. K. (2018). Normal observers show no evidence for blindsight in facial emotion perception. In *bioRxiv* (p. 314906). https://doi.org/10.1101/314906

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, *11*(1), 2634.

Rosenthal, D. (2019). Consciousness and confidence. *Neuropsychologia*, *128*, 255–265.

Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175.

Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating Perceptual Confidence from Discrimination Accuracy Reveals No Influence of Metacognitive Awareness on Working Memory. *Frontiers in Psychology*, *7*, 851.

Samaha, J., Iemi, L., & Postle, B. R. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2017.02.005

Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *biorxiv;369140v2*, 590.

Samuelson, P. A. (1942). A Note on Alternative Regressions. *Econometrica: Journal of the Econometric Society*, *10*(1), 80–83.

Sherman, M. T., Barrett, A. B., & Kanai, R. (2015). Inferences about consciousness using subjective reports of confidence. *Behavioral Methods in Consciousness Research*, 87–106.

Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception & Psychophysics*, *77*(1), 258–271.

Stolyarova, A., Rakhshan, M., Hart, E. E., O'Dell, T. J., Peters, M. A. K., Lau, H., Soltani, A., & Izquierdo, A. (2019). Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. *Nature Communications*, *10*(1), 4704.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. https://philarchive.org/rec/WEIBAC

Zehetleitner, M., & Rausch, M. (2013). Being confident without seeing: what subjective measures of visual consciousness are about. *Attention, Perception & Psychophysics*, *75*(7), 1406–1426.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.

# The metaperceptual function: Exploring dissociations between confidence and task performance with type 2 psychometric curves

Brian Maniscalco[1*†], Olenka Graham Castaneda[2*], Brian Odegaard[3], Jorge Morales[4], Sivananda Rajananda[2], Megan A. K. Peters[1,2]

## Supplementary Material

## S1. Group-level fits from averaged single-subject fits

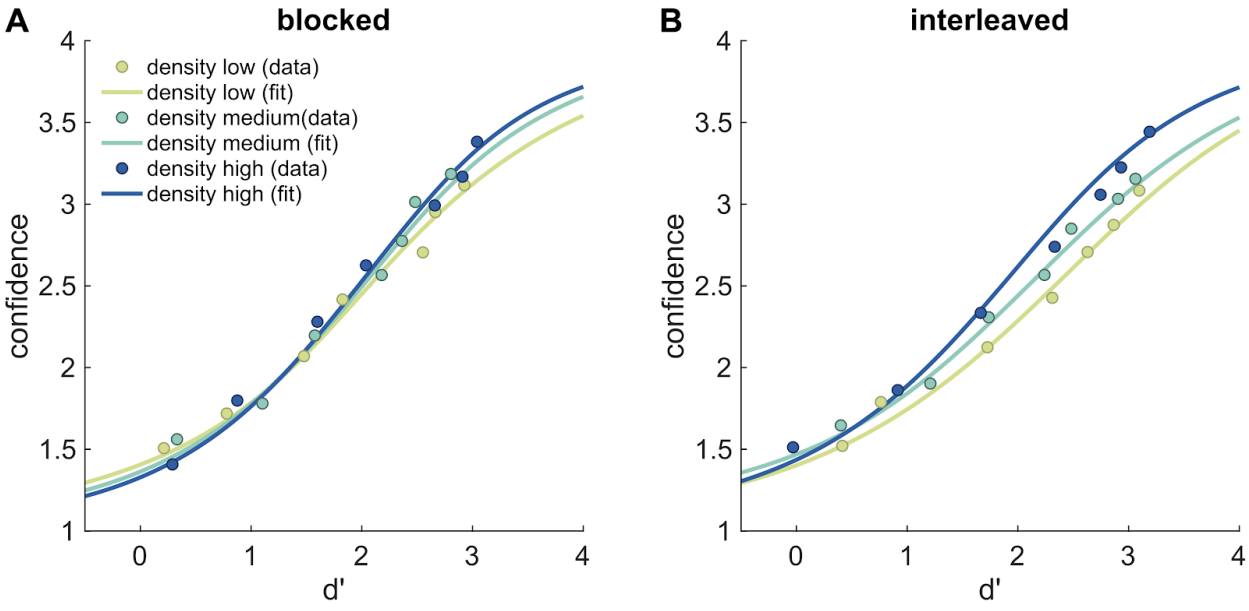


**Figure S1.** Metaperceptual curve fits to group-averaged confidence and d' data obtained by averaging metaperceptual curves fitted to single-subject data. These are similar to the fit obtained by fitting directly to the group-averaged confidence and d' data, as shown in Figure 3.

## S2. Alternative approaches to fitting the type 2 psychometric function

In the interest of comprehensiveness, we also performed another version of the metaperceptual curve fitting analysis presented in the main text (Section 3.1) in which where there was an additional parameter, $c_b$, that determined minimum confidence of the logistic function fit, i.e. instead of confidence falling in range [1,4], confidence ranged over [$c_b$,4]. The rationale for this change is that for most participants, confidence > 1 even when d' = 0, so the data might be better modeled if allowed to have a confidence floor > 1.

In this case, the modified equations become

$$conf = f(d' \mid \mu,\ s)\ = (4 - c_b)\ \left(\frac{1}{1+e^{-(d'-\mu)/s}}\right) + c_b \qquad \text{(S1)}$$

$$d' = f^{-1}(conf \mid \mu,\ s)\ =\ -s\ log\left(\frac{4-conf}{conf-c_b}\right) + \mu \qquad \text{(S2)}$$

where $c_b$ is the baseline or floor level of confidence. Note the equations used in the main analysis are a special case where the value of $c_b$ is fixed at 1. Also note that these equations require constraining $c_b$ to be less than the smallest confidence value in the data set to be fitted (i.e. $c_b < min(conf)$), otherwise the results are infinite (when $conf = c_b$) or imaginary (when $conf < c_b$).

Allowing $c_b$ to be a free parameter produces comparable metaperceptual function fits to the data as those produced in the main analyses in which $c_b$ is fixed at 1 (compare Figure 3 with Figure S2, Figure S1 with Figure S3, and Figure 4 with Figure S4), and the extra free parameter also leads to slightly lower overall error in the data fits. When reproducing the dot density x block type ANOVA analysis on single-subject fits to the μ parameter derived from this method, the main effect of dot density on μ is still robust ($F(2,40) = 8.72$, $p = 7e\text{-}4$), but the dot density x block type interaction becomes slightly weaker than in the fits with fixed $c_b$ reported in the main text ($F(2,40) = 2.97$, $p = 0.063$). In post-hoc ANOVAs investigating the effect of dot density on μ separately for the Blocked and Interleaved conditions, it remains the case that the main effect of dot density is significant in the Interleaved ($F(2,40) = 9.18$, $p = 5e\text{-}4$) but not the Blocked ($F(2,40) = 1.73$, $p = 0.2$) condition.

However, with $c_b$ as a free parameter, some single-subject fits appear highly implausible due to confidence remaining at floor over large ranges of d', only to rise suddenly at d' values well above threshold (e.g. confidence at floor for d' < 2, with sharp increases for d' > 2 ). Although this behavior can yield lower-error fits in noisy data, it is theoretically implausible. We show a few single-subject examples of this situation in Figure S5 (left column), alongside the

corresponding fits without the $c_b$ parameter (right column); for the latter, the fitting error is higher but the fits are much more conceptually plausible. Thus, in the main analyses we do not fit $c_b$ as a free parameter, because (1) it turns out that even with $c_b$ fixed at 1, the model fits can capture well the phenomenon whereby confidence > 1 when d' = 0; and (2) some of the single-subject fits achieved with $c_b$ as a free parameter achieve lower error only through highly implausible fits.

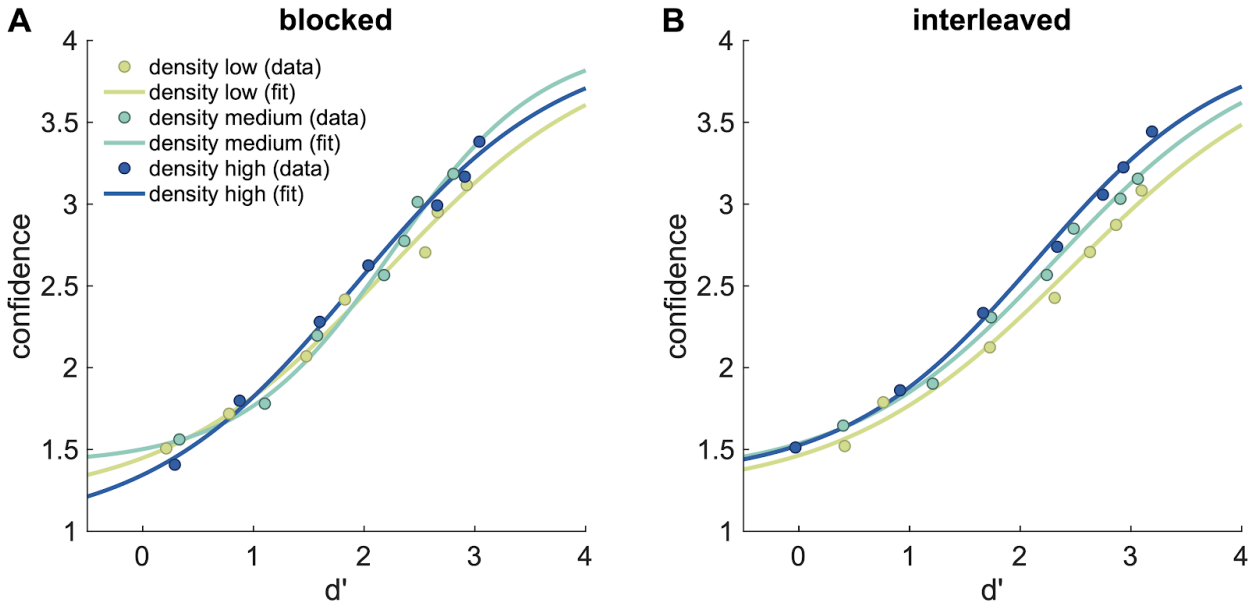


**Figure S2.** Metaperceptual curve fits obtained by fitting directly to group-averaged confidence and d' data, with $c_b$ included as a free parameter in the curve fit. Compare to fit using $c_b$ fixed at 1 as presented in Figure 3.

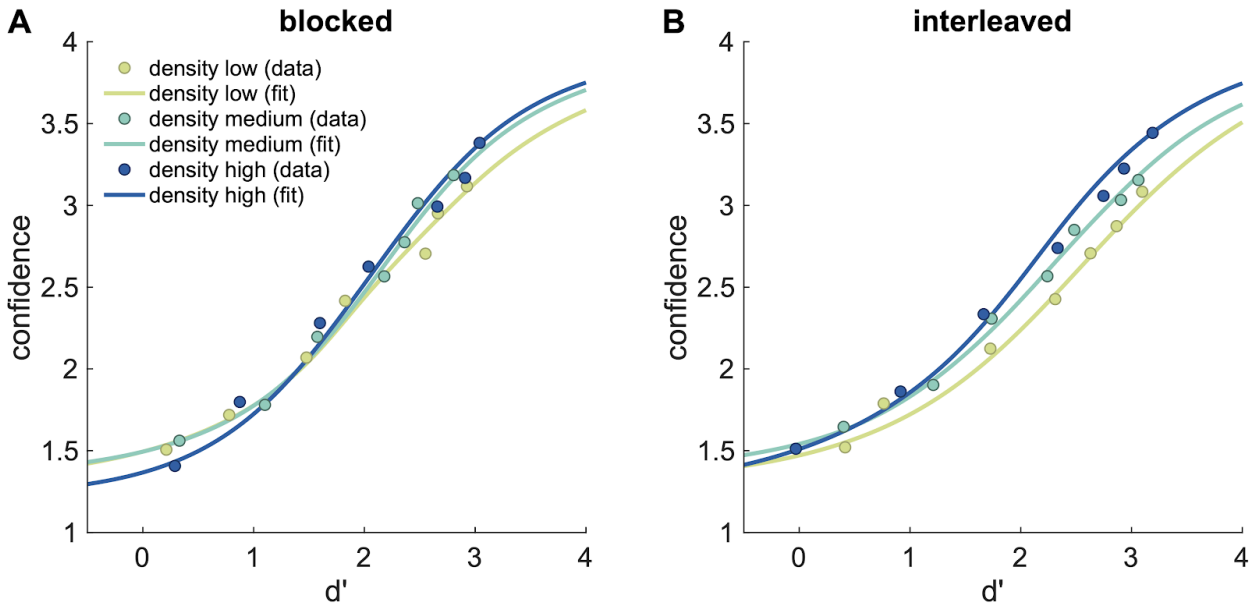


**Figure S3.** Metaperceptual curve fits to group-averaged confidence and d' data obtained by averaging metaperceptual curves fitted to single-subject data, with $c_b$ included as a free parameter in the curve fit. Compare to fit using $c_b$ fixed at 1 as presented in Figure S1.

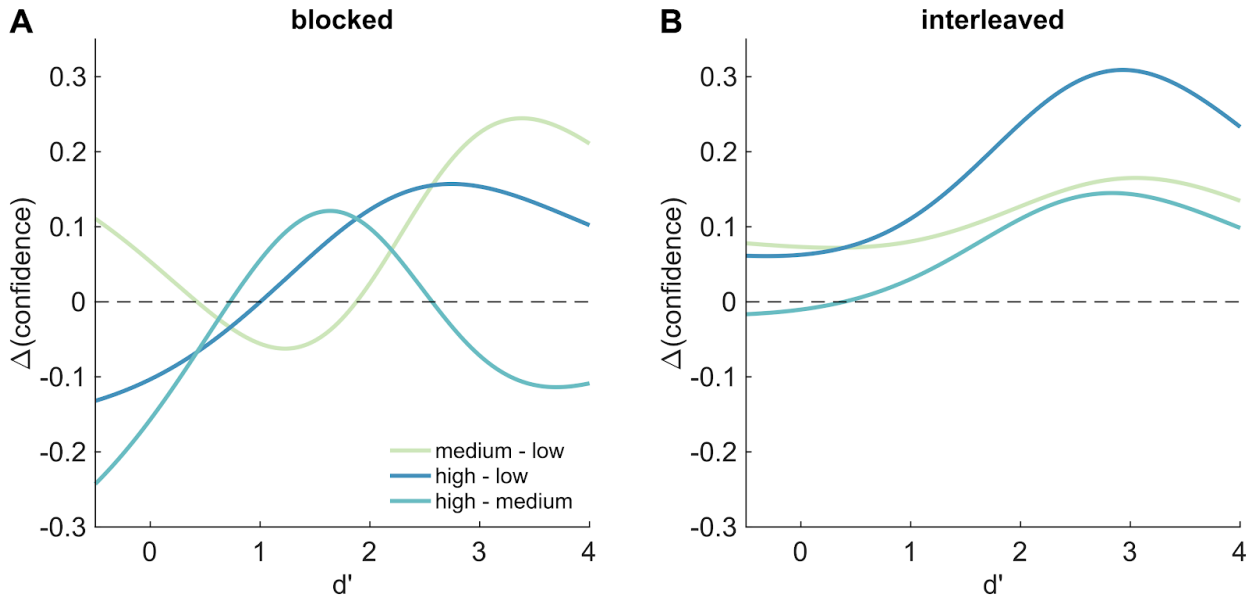


**Figure S4.** Difference curves for metaperceptual curves fitted to group-average data, with $c_b$ included as a free parameter in the curve fit. Compare to fit using $c_b$ fixed at 1 as presented in Figure 4.

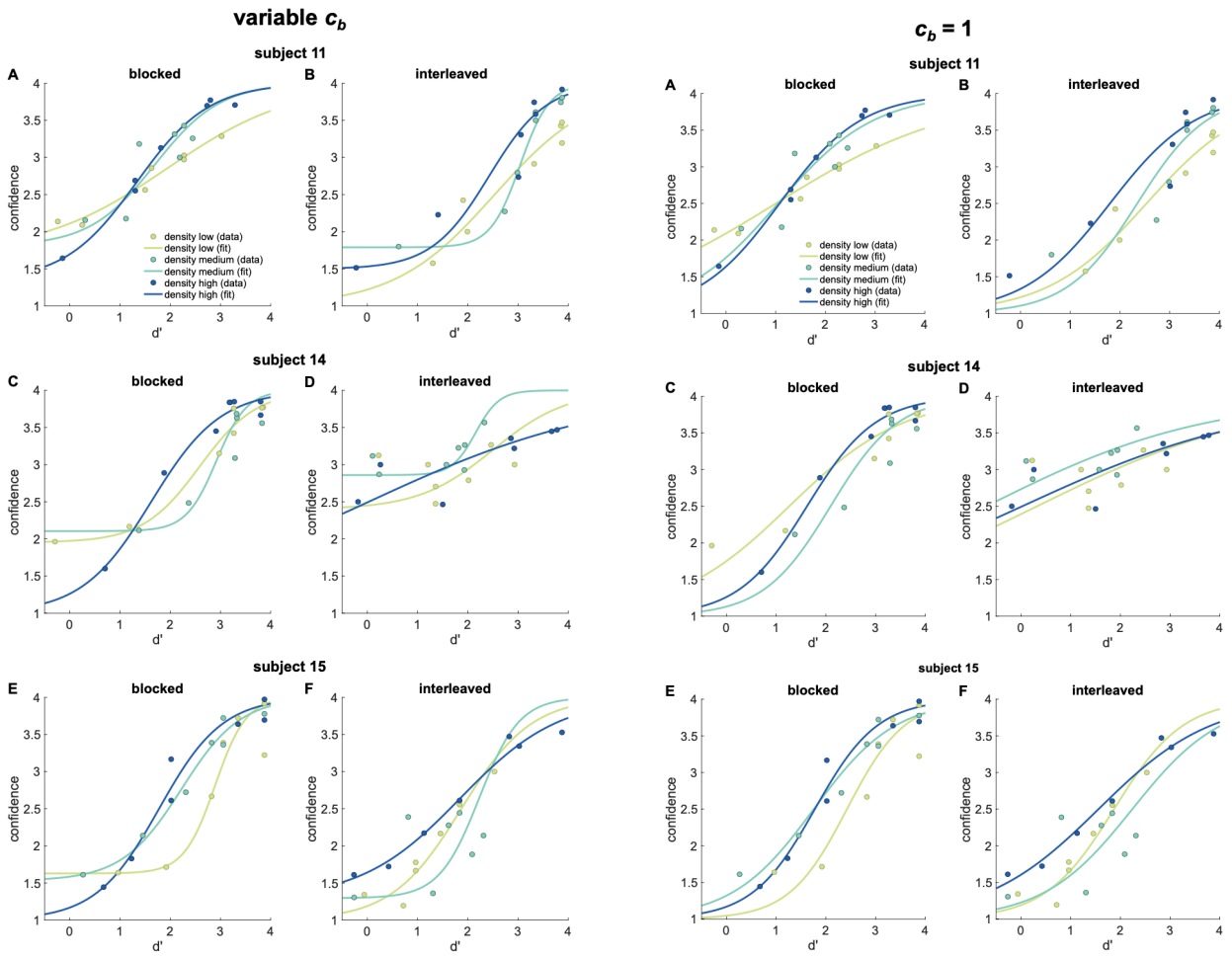


**Figure S5.** Example single-subject metaperceptual function fits for Blocked and Interleaved conditions when $c_b$ is allowed to vary (two leftmost columns) instead of being fixed at $c_b = 1$ as in the main analysis (two rightmost columns). Single-subject fits were chosen to highlight implausible fitting behavior when $c_b$ is left as a free parameter: note that when $c_b$ is fitted as a free parameter, some fitted curves exhibit floor levels of confidence up until d' ≅ 2, and then abruptly rise in confidence when d' > ~2. By contrast, the fits achieved with $c_b$ fixed at 1 exhibit more conceptually plausible behavior across the entire d' range, in spite of exhibiting slightly higher fitting error.

## S3. Connections between Sections Sections 3.1 and 3.4

We performed an additional analysis to further examine the differences for metaperceptual curve fits for Low, Medium, and High dot density conditions, but this time with data points plotted as well (Figure S6, following Figure 4 in the main text). Data are plotted from all conditions where the main effect of dot density on d' has p > 0.1. The value for d' used is the average d' across all three densities. The vertical gray shaded regions are plotted to indicate data points

coming from the same motion coherence condition and having similar (i.e, not significantly different) d'.

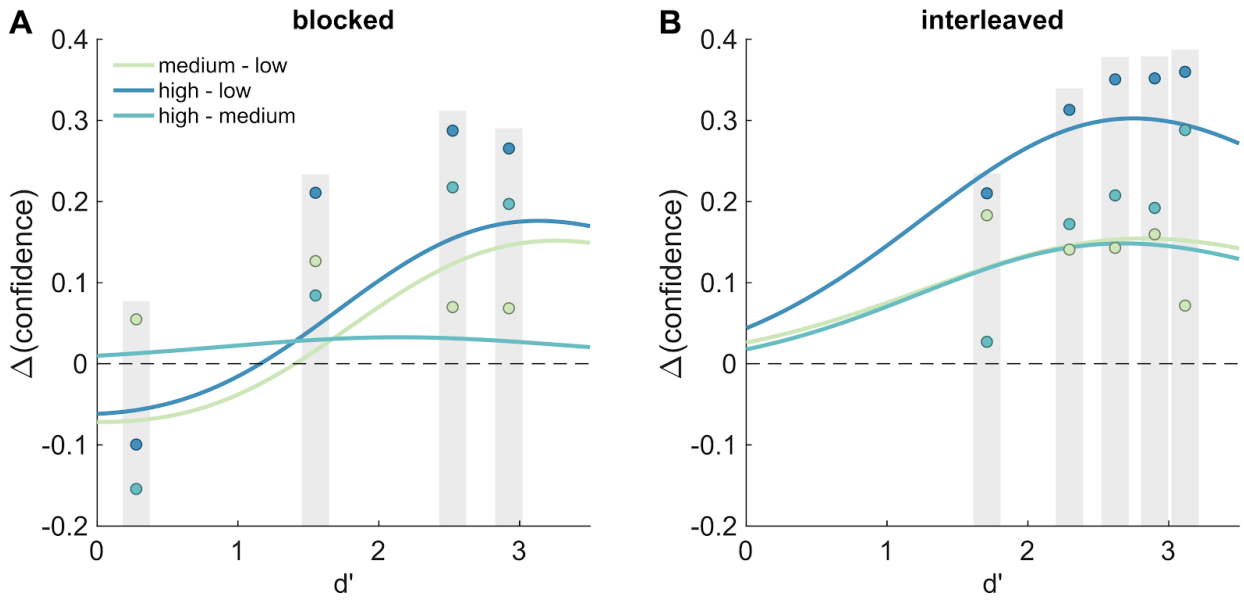


**Figure S6.** Difference curves for metaperceptual curves fitted to group-average data. The group-average fits shown here are identical to those shown in Figure 4 of the main manuscript. The circles within each gray shaded band indicate group averages from the same motion coherence condition that have similar d'.

## S4. Accounting for the effect of dot density on confidence in a signal detection theory framework

We conducted a simple signal detection theory (SDT) simulation to illustrate the idea that higher dot density might be associated with higher variability in perceptual evidence, which in turn could lead to higher reports of perceptual confidence ((Macmillan & Creelman, 2004); see also (Morales et al., 2019) for more extensive discussion of the technical details of how this effect is modeled in SDT). In the simulation, we made the following assumptions: (1) increasing motion coherence leads to increasing distance between the means of the evidence distributions; (2) increasing dot density leads to increasing variance in the evidence distributions; (3) decision criteria used to make perceptual decisions and rate confidence are constant across conditions. In order to yield a reasonable range of confidence values across levels of d', we chose the location of the confidence criteria so as to yield a roughly uniform probability of rating confidence as 1, 2, 3, or 4 when d' in the Medium dot density condition had an intermediate value of 1.5. We set the standard deviations of the evidence distributions for Low, Medium, and High dot density conditions to 0.8, 1, and 1.2. With these values fixed, we swept the distance

between the evidence distributions from 0 to 4.8 in step sizes of 0.1, computing at each of these values the corresponding d' and mean confidence.

Results of this simulation are plotted in Figure S7a. As expected, confidence increases monotonically with d', and conditions with higher evidence variance have higher mean confidence, consistent with the hypothesis that higher dot density is associated with higher variability in perceptual evidence, which in turn leads to higher confidence.

However, this simple model fails to capture some features of the data (Figure 3), most notably the fact that the effect of dot density on confidence increases with increasing d' (Figure 4), rather than remaining roughly constant across all d' values (Figure S7a). Additionally, the slope of the metaperceptual curve exhibited over the full range of d' values is too shallow in the simulation (confidence ranging from about 2 to 3.5 over the region $0 \leq d' \leq 3$) as compared to the data (confidence ranging from about 1.5 to 3.5 over the region $0 \leq d' \leq 3$). Finally, although not evident in Figure S7a, the simple model also makes the incorrect prediction that d' should decrease with increasing dot density within a given motion coherence level (since motion coherence controls distance between the evidence distributions, dot density control evidence standard deviation, and d' = (distance between evidence distributions) / (standard deviation of evidence distributions)). Thus, although this simple SDT model may give some preliminary insight on possible mechanisms underlying the observed effects, it cannot be the whole story.

We conducted a second SDT simulation, this time based on a two-dimensional formulation which allows for modeling of the positive evidence (PE) decision rule for confidence, a phenomenon whereby subjects tend to base perceptual confidence only on evidence that is consistent with their perceptual decision ("positive evidence") while neglecting to take into account conflicting evidence (Maniscalco et al., 2016; Odegaard et al., 2018; Peters, Thesen, et al., 2017; Zylberberg et al., 2012); see (Morales et al., 2019) and (Maniscalco et al., 2016) for more extensive discussion of the technical details of how this effect is modeled in 2D SDT. Simulation procedure for the 2D SDT modeling proceeded similarly to the process described above for the simpler 1D SDT model.

Results of the 2D SDT simulations are presented in Figure S7b. The 2D SDT model was similarly able to capture the effect of density on confidence, while also having a steeper overall slope for the confidence vs d' curve, in closer agreement to the slopes observed in the empirical data. However, the simple 2D SDT model was similarly unable to capture the effect of increasing differences in confidence with increasing d', and similarly made the incorrect prediction that for a fixed motion coherence level, d' should decrease with increasing dot density.

Note that we intend these model simulations to be of heuristic and illustrative value only; it is possible that more complex formulations of these models could capture more features of the data set.

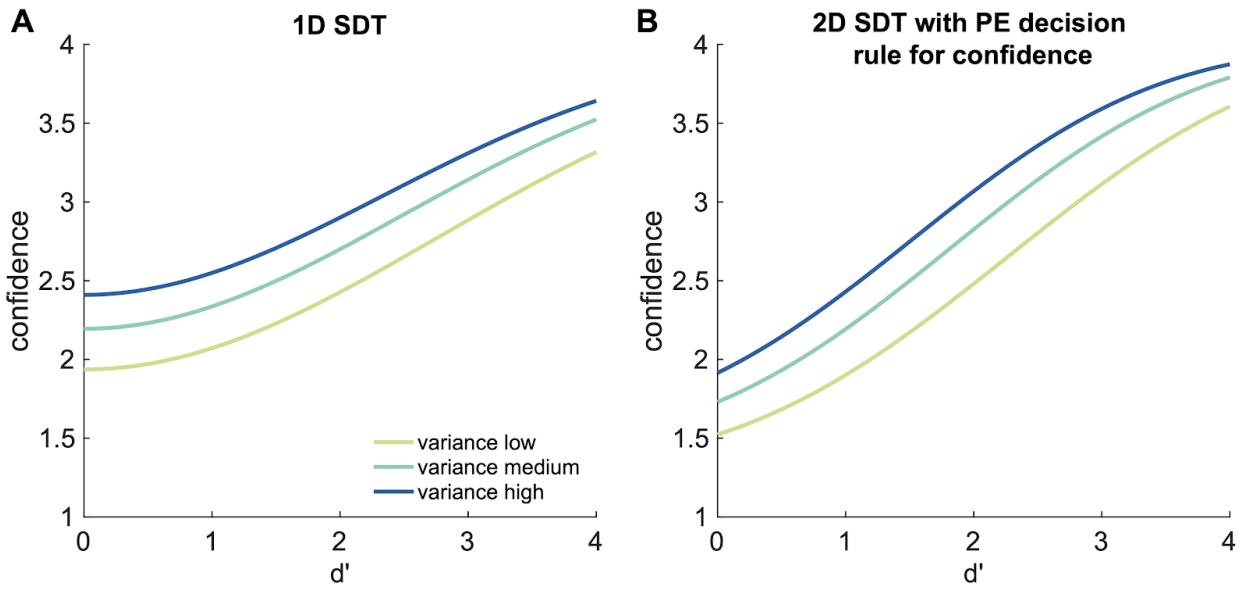


**Figure S7.** Metaperceptual curves resulting from simple 1D and 2D SDT model simulations. (A) One-dimensional SDT can qualitatively capture the MPDC effect, but slopes of the metaperceptual functions are too shallow and across-condition differences in confidence are too large at low d' values relative to empirical data. (B) 2D SDT with the PE confidence rule (Maniscalco et al., 2016; Miyoshi & Lau, 2020; Peters, Thesen, et al., 2017) produces a slope that better matches the empirical data presented in the main text, although it still overestimates across-condition differences in confidence at low values of d'.

## S5. Online pilot data

Prior to collecting the laboratory data described in the main text, we conducted pilot experiments on Amazon Mechanical Turk using experimental designs similar to the one used for the data set reported in the main manuscript in order to achieve a preliminary assessment of the potential effects of motion coherence, dot density, and block type on d' and confidence.

*Experimental design*

Unlike the main data set, here the block type manipulation (Blocked vs Interleaved) was a between-subject, rather than within-subject, factor. There were five levels of motion coherence evenly spaced between 0.2 and 1, and three levels of dot density for which 200, 600, or 1000 dots in total were presented in the RDK display. The rectangular aperture which contained all the dots was 800 pixels wide x 400 pixels high and positioned in the center of the screen. The fixation cross was at the center of this rectangle. The 2 circular regions in which coherent motion could occur had diameters of 300 pixels and were centered 200 pixels to the left and right sides of the fixation cross. Each dot had a diameter of 4 pixels. Dot speed was 1 pixel /

frame in the Interleaved condition; this was increased to 2 pixels / frame in the Blocked condition in an effort to boost task performance, due to many participants exhibiting near chance-level performance in the Interleaved condition. Each participant completed 300 trials total, yielding 20 trials for each level of motion coherence x dot density. Due to a programming error, trial counts were not perfectly uniformly distributed across conditions for the Interleaved data. Additionally, for the first 3 of 10 participants in the final sample submitted to analysis in the Interleaved condition, the region of coherent motion was smaller than it was for other participants, and for these participants coherent motion moved left or right rather than moving downwards. This was subsequently changed to downward motion in order to eliminate any potential response conflict in cases where motion direction (left / right) conflicted with location of the region of coherent motion relative to fixation (left / right). Code for the online task can be found at https://github.com/vrsivananda/MPDC_RDKpsychometric.

*Participants*

In total, 33 participants completed the Interleaved condition. 6 of these were lab members and the remaining 27 were recruited via Amazon Mechanical Turk. Only 9 of 27 Mechanical Turk participants had data suitable for analysis (i.e. mean task performance above chance levels and full usage of the confidence rating scale), leaving 15 participants total. However, a programming error causing uneven distribution of trial counts across dot density conditions led to the loss of 5 further participants, leaving n=10 for the final sample. Of this n=10 sample, 4 participants were lab members, and 3 of these were aware of the hypothesis linking confidence with dot density.

In total, 19 participants completed the Blocked condition. 3 of these were lab members and the remaining 16 were recruited via Amazon Mechanical Turk. Only 6 of 16 Mechanical Turk participants had data suitable for analysis (i.e. mean task performance above chance levels and full usage of the confidence rating scale), leaving 9 participants total.

As with the main study, all participants provided consent to participate in the study (by clicking an “I agree” consent box in the web interface) and all procedures were approved by the University of California Riverside Institutional Review Board.

*Analysis*

In Figure S8, we plot mean confidence and d’ as a function of block type, motion coherence, and dot density for the online experiment participants with usable data. As in Figure 3, we also plot logistic function fits to the group-averaged data. In spite of numerous suboptimal features of this data set -- high data quality attrition rate leading to small sample size, relatively low trial counts for each participant, minor differences in design across participants, etc -- we still observed a pattern of results qualitatively similar to those reported in the main data set (Figure 3). In particular, higher dot density was associated with higher confidence across a broad range of d’ values, and this effect appeared more pronounced when density was interleaved rather than blocked. Importantly, the qualitative effect of dot density on confidence in the interleaved condition (Figure S8b) remained even after the 3 participants who were not naive to the main hypothesis that dot density correlates with confidence were omitted.

For the sake of completeness, we performed metaperceptual function fits to single-subject confidence vs d' curves and submitted the logistic parameter μ to a mixed-design block type x dot density ANOVA. Not surprisingly due to low statistical power and noisy data, this analysis revealed only a marginal main effect of dot density ($F(2,34) = 2.81$, $p = 0.074$), and a non-significant block type x dot density interaction ($F(2,34) = 0.85$, $p = 0.4$). Nonetheless, the qualitative patterns in the data are consistent with those observed in the more rigorous laboratory sample and thus constitute a modest, qualitative replication of the main findings.

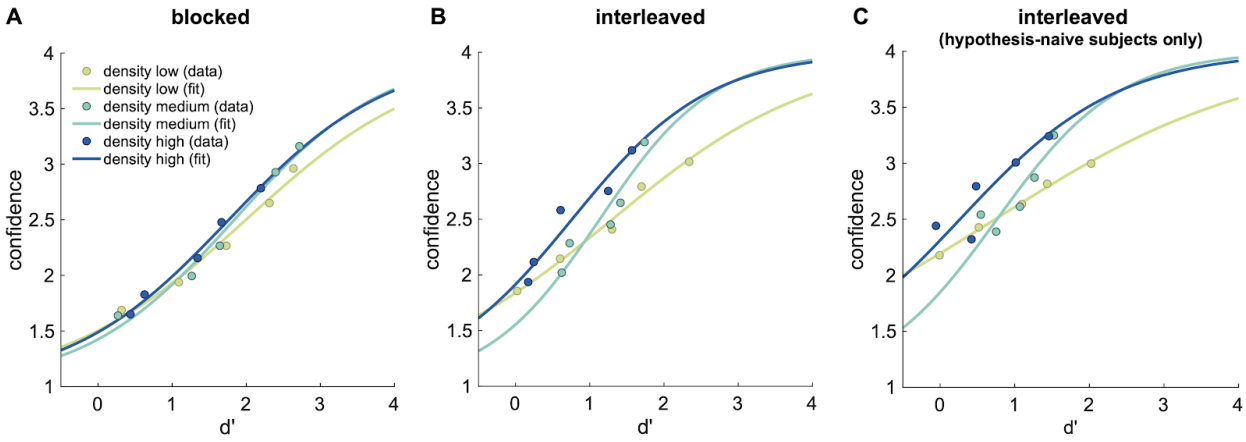


**Figure S8.** Data and metaperceptual function fits from Amazon Mechanical Turk pilot experiment. (A & B) Results were qualitatively similar to those observed in the main data set (Figure 3). (C) The qualitative effect of dot density on confidence (see panel B; n=10) remains even when 3 participants who were not naive to the hypothesis that confidence correlates with density were removed from the analysis (n=7).