Taylor & Francis
Taylor & Francis Group

Theoretical Issues in Ergonomics Science

ISSN: 1463-922X (Print) 1464-536X (Online) Journal homepage: https://www.tandfonline.com/loi/ttie20

Model of variability estimation: factors influencing human prediction and estimation of variability in continuous information

Christopher D. Wickens, Benjamin A. Clegg, Jessica K. Witt, C. A. P. Smith, Nathan Herdener & Kimberly S. Spahr

To cite this article: Christopher D. Wickens, Benjamin A. Clegg, Jessica K. Witt, C. A. P. Smith, Nathan Herdener & Kimberly S. Spahr (2019): Model of variability estimation: factors influencing human prediction and estimation of variability in continuous information, Theoretical Issues in Ergonomics Science, DOI: 10.1080/1463922X.2019.1679907

To link to this article: https://doi.org/10.1080/1463922X.2019.1679907

Published online: 04 Nov 2019.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ttie20

THEORETICAL ISSUES IN ERGONOMICS SCIENCE
https://doi.org/10.1080/1463922X.2019.1679907

Taylor & Francis
Taylor & Francis Group

Check for updates

Model of variability estimation: factors influencing human prediction and estimation of variability in continuous information

Christopher D. Wickens[a], Benjamin A. Clegg[a], Jessica K. Witt[a], C. A. P. Smith[b], Nathan Herdener[a] and Kimberly S. Spahr[a]

[a]Department of Psychology, Colorado State University, Fort Collins, CO, USA; [b]Computer Information Systems Department, Colorado State University, Fort Collins, CO, USA

ABSTRACT
Understanding the variability of trends and other continuously distributed quantities is a vital ability underlying many safety critical decisions, such as how widely to search for a downed aircraft, or whether to prepare for evacuation in the face of an uncertain hurricane or hurricane track. We first review the sparse research on this topic which indicates a general systematic tendency to underestimate such variability, akin to overconfidence in the precision of prediction. However, the magnitude of such underestimation varies across experiments and research paradigms. Based on these existing findings, and other known biases and vulnerabilities of the perception and cognition of multiple instances, we define the core elements of a computational model that can itself predict three measures of performance in variability estimation: bias (to over or underestimate variability), sensitivity (to variability differences) and precision (of variability judgements). Factors and approximate weighting in influencing these measures are then identified regarding attention, the number of instances across whose variability is estimated, the time delay affecting the memory system employed, familiarity of material, the anchoring heuristic and the method of judgement. These are then incorporated into foundations for a linear additive model.

ARTICLE HISTORY
Received 23 October 2018
Accepted 9 October 2019

KEYWORDS
Prediction; trajectories; decision making; model; variability; judgement; bias

1. Relevance to human factors/Relevance to ergonomics theory

In many real-world environments, such as those involving severe storm alerting, or process control, people must make safety-critical decisions based upon their prediction or extrapolation of trending quantities. If they underestimate the potential variability or uncertainty in such predictive trends and fail to prepare, for example for worst case scenarios, the consequences can be severe. This paper reviews scarce literature on this process, concludes that there is a general tendency to underestimate variability and then presents the framework

CONTACT Christopher D. Wickens pandawickens94@aol.com Department of Psychology, Colorado State University, Fort Collins, CO, USA.
© 2019 Informa UK Limited, trading as Taylor & Francis Group

and key coefficients of an information processing model that is designed to predict the magnitude of this variability underestimation bias.

2. Expressions of variability in judgement

Variance is all around us, from the increasing variability of climate patterns to the vast diversity of races and religions in the US to the wide range of human performance (Muhs, Karwowski, and Kern 2018) to the very ANOVA that we use to analyse data. Correctly comprehending the potential variability is often essential to effective performance in many scenarios. This paper examines people's understanding of the variance and uncertainty inherent in *continuous trends over time*; for example, the trend of election polling, of the stock market or other economic indicators, the movement of a severe storm, the change in global climate (Ray et al. 2015; McCarthy et al. 2015), the change in patient health indicators, a process control variable or an aircraft trajectory. Such trends have two important features. First, the variables underlying these trends are *continuously distributed* or analogue in form, not the discrete events such as a gamble, an earthquake or election outcome (which have received the focus of most research in decision science of uncertainty). Second, a primary purpose for the user is often to understand *the prediction, or extrapolation* of the trend. Furthermore, the prediction itself can be broken down into two elements: where is the trend most likely to go; its *mean* or central tendency, and how certain are we of this mean; that is, what is its judged *variability?* It is this latter commodity, variability, that is of greatest interest in our current analysis.

To preview what lies ahead in this paper, one of the phenomena revealed by our own research and those of others employing related paradigms is that people are poor at understanding potential variability. In particular, there is a general tendency to *underestimate the amount of variability*, whether of predicted trends or other continuously distributed quantities. Some researchers have referred to this underestimation of variability as overconfidence – greater confidence than is warranted in the precision of estimating or predicting the mean. There is a broad existing literature on overconfidence in a variety of realms of decision and judgement research (e.g. Einhorn and Hogarth 1978; Fischoff, Slovic, and Lichtenstein 1978; Juslin, Winman, and Hansson 2007; Kahneman 2011; Moore and Healy 2008; Olsson 2014; Tetlock 2005). This research, while perhaps related is not the focus of this paper. Instead, we examine the behavioural expressions of the accuracy and bias in judging variability of *continuously distributed quantities*.

The interpretation of variability is of considerable ergonomic importance because of the consequence of decisions that flow from such judgements. To the extent that such variability is underestimated, poor choices may be made. For example, people who underestimate the potential variability of a path of a hurricane and are outside the centre of the predicted path, will be more reluctant to evacuate than they should be. As another example, the process control monitor who underestimates the possible upper boundary of an increasing trend (e.g. of pressure) may fail to prepare adequately for an out of bounds catastrophic event (Strobhar 2014). As a third example, in search and rescue, those who are overconfident of their prediction of where a downed aircraft may be, may inappropriately narrow the radius of their search.

In this paper, we first review of the relevant literature, and because we identify so little literature that directly examines understanding variability of predicted trends, we expand this review to consider a broader range of research that has examined biases in judged variability of continuously distributed quantities, whether in prediction of future trends or of assessment of current state. This offers a foundation to understand the factors that contribute to biases, or inversely, calibration in variability estimation. Following this review, we describe three quantitative metrics of variability estimation, and present a perceptual-cognitive model of variability estimation (MOVE). The model considers causal mechanisms that were revealed in the literature. We note that in this paper we do not address the perceptual aspects of rapid trend prediction, such as predicting the flight of a ball in sports (Suss and Ward 2015) or the trajectory of an automobile (Engström et al. 2018), or the judgement of time to contact in collisions (DeLucia 2015). Such research, of critical importance, does not generally address the uncertainty or variability in those trajectories, and the more cognitive aspects of task performance.

3. Review of the literature

One approach that speaks to the understanding of variability are studies concerning humans as 'intuitive statisticians' (Peterson and Beach 1967; Pollard 1984), some of which have examined how well people estimate variance. For example, Beach and Scopp (1968) inferred people's estimation of variance of sets of digits, across different sets that differed in their true variance. They observed that people significantly underestimated variance, particularly when that variance was large. Lathrop (1967) also had people make intuitive estimations of the variance in sets of lines, observing that such estimations were smaller with larger means, but without any reported results related to the calibration of judgements. Pollard (1984) summarised these and other findings (Levin et al. 1977; Lovie 1978; Lovie and Lovie 1976) to conclude that people are not very proficient at this task, but that the methodologies of some of these studies are flawed.

Obrecht, Chapman, and Gelman (2007) asked people to do 'intuitive *t*-tests' of differences and found that people greatly under-estimated the contributions of variance (relative to the mean difference and sample size) to the significance of differences, consistent with under-estimating the magnitude of the former. In contrast, Pitz (1980) offers a model of intuitive variance estimation that speculates people focus excessive attention on outliers of a distribution (Hamilos and Pitz 1977). If this were the case, it would suggest an overestimation of variance. However, Pitz does not report empirical variance estimation data to support this supposition.

Although not expressly framed within the domain of intuitive statistics, two studies do bear directly on the calibration of variability estimation. Kareev, Arnon, and Horwitz-Zeliger (2002) assessed participants' judgements of variability of spatial stimuli (e.g. coloured cylinders, coloured matches). Using a variety of methodologies, across five different experiments, they found evidence for underestimation in these spatial analogue quantities, even as they observed that people were sensitive to differences in variability between sets. They propose a model that explains the underestimation effect from an inherent statistical bias in estimating population variance from smaller samples. This is associated with a limited capacity to hold samples in working memory, and they observed that those with smaller working memory capacity showed a larger under-estimation bias.

Hansson, Juslin, and Winman (2008) examined bias in estimating the variability of a set of continuously distributed hypothetical company earnings. They found significant underestimation of the true variability. Importantly for the model we present below, these findings were strongest, most consistent, and least influenced by training and practice, when participants used a method of adjusting a subjective confidence interval to estimate variability. As with Kareev, Arnon, and Horwitz-Zeliger (2002), Hansson, Juslin, and Winman (2008) also found an inverse relationship of bias with working memory capacity.

A different, non-laboratory example of underestimating the variability of analogue quantities is provided by Henrion and Fischhoff (1986), who found that scientists tended to under-estimate the amount of variability in measurements of scientific constants under their scrutiny, such as the speed of light (*c*), particle masses, and the proton's magnetic moment. At one time, scientists discovered these phenomena, but with overconfidence, their published results did not consider scientific uncertainty. Closely related, Soll and Klayman (2004) found that people often set too narrow confidence intervals around estimated values.

In our own laboratory, we have examined peoples' ability to predict the mean and variability of a set of trajectories, not unlike the set of hurricane tracks shown on the left of Figure 1 (Herdener et al. 2016, 2018, 2019a, 2019b; Pugh et al. 2018). The paradigm is illustrated on the right side of Figure 1. In this paradigm, over a time period of about 30 s, people are shown a set of trajectories, with a start point time (T0) and an early point in the trajectory (T1) depicted. They are then are asked to predict the location at T3, the endpoint, given not only the two previously seen points but also their memory or 'mental model' of the typical path and variability of *all* tracks seen previously. At the end of a block of estimating the trajectory endpoints, they are then asked to estimate the variability of the tracks they had seen, by estimating how many of them terminated within a circle of a given diameter. Overwhelmingly, people underestimate this variability, by as much as 30%, by estimating far more terminations within the circle than they had actually seen. Such

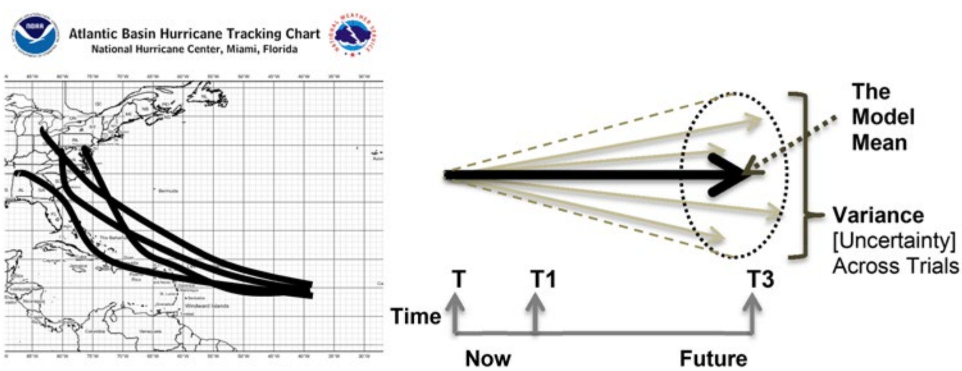

Figure 1. Examples of trajectory variability. Left: spaghetti plot hurricane trajectories. Right. Generic trajectories from Colorado State University research. Each line represents a trajectory, presented sequentially, whose endpoint is predicted. After all instances have been presented, participants judge the variability of those they had seen by estimating how many terminated in a circle of a fixed radius.

underestimation extends to predicting uncertain ship trajectories as well (Wickens et al. 2019). It should be noted however that within our own lab, the findings of underestimation of variability is not universal (Spahr et al. 2018; Spahr 2019), approaching calibration and even overestimation. In presenting our model, we describe some of the reasons for discrepant results.

Another contrary example has been found with studies of ensemble perception, for which participants view a group of objects and make a judgement about a summary statistic such as the mean colour or the variability of their orientations. Most of these studies focused on perception of the mean (Whitney and Yamanashi Leib 2018). In a study to examine accuracy and biases in perception of the variability (Witt 2019), participants viewed a sequence of nine lines. The lines differed in their orientations from a small amount (1° between lines) or a large amount (8° between lines), and participants judged whether the spread was small or large. They showed a 50% *overestimation* of the variability of the line orientations (Witt 2019). This suggests that memory demands (present in the trajectory studies above, but absent in ensemble perception) may play a strong role in the under-estimation tendencies observed.

4. Quantifying variability estimation

Measures of variability judgement, described in detail can yield three different measures of performance, as depicted in Figure 2, which plots judged variability against true variability. The 45-degree diagonal dotted grey line in each panel represents a line of perfect calibration where estimated variability matches true variability. Any estimations in the upper quadrant above the dotted line represent overestimation of variability, and those in the lower quadrant represent underestimation.

If multiple levels of true variability are assessed, then:

The measure of *sensitivity* (see left panel of Figure 2) can be determined from the slope of the variation in estimated variability across true variability. For perfect sensitivity every increase in true variability produces an identical increase in estimated sensitivity (slope of 1.0). No sensitivity or calibration (slope = 0), is reflected by the flat thin solid line where changes in true variability are not reflected in estimate differences: the human is

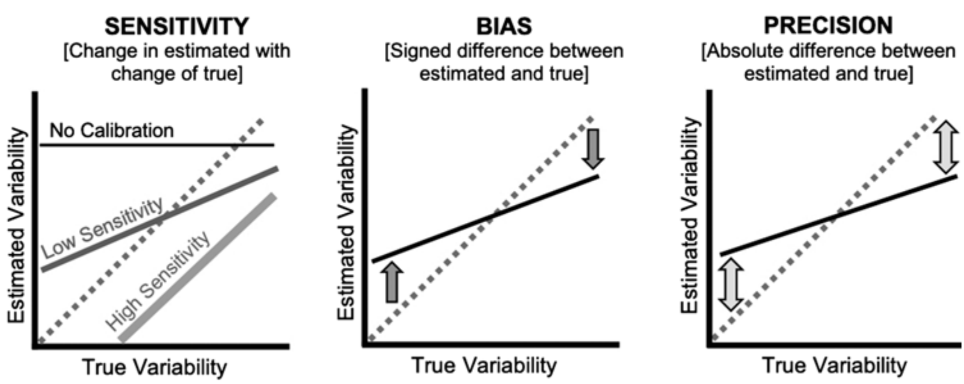

Figure 2. Three possible measures of performance in the relation between actual or true variability of a distribution and perceived or judged (i.e. subjective) variability.

'variability-blind'. The steep, thick line represents high sensitivity, with good tracking of changes in true variability in the person's estimate; Note that even though this line does not align perfectly with the dashed line, it is still representative of high sensitivity. The difference in the intercept between the two line corresponds to bias (see below). The 'low sensitivity' line represents diminished sensitivity relative to the optimal, with some change in reported variability as true variability increases. Sensitivity can be defined as the ratio of slopes of an obtained function to the optimal, grey dotted line. Beach and Scopp (1968) present a corresponding graph, indicating some, but reduced, sensitivity to variability differences.

The measure of *bias* towards over- or underestimation of variability can be derived at any single level of true variability and is simply the *signed difference* of estimated minus true (as illustrated by the vertical, grey arrows in Figure 2 middle panel, for the left arrow the estimated variability is greater than the true and for the other arrow the estimated is lower than the true variability). Thus, greater positive numbers indicate more over-estimation of variability and negative numbers indicate under-estimation of variability.

The measure of *precision* can be derived by the *absolute difference* between estimated and true variability at any single level of true variability (as illustrated by the vertical double headed arrows in Figure 2 right panel, which at each of the two levels of true variability both match in precision). The distinction between precision and bias is important when group data are analysed. For example, while the mean value of bias may be 0, this can result from a group that has low precision, if judgements are equally represented by people with very positive and very negative bias.

In applying the representation above, to people's judgements of variability, we can assume that sensitivity and bias can operate independently, such as is often the case with signal detection theory (Wickens 2002). Note that the high sensitivity line in the left panel in Figure 2 shows good sensitivity as the slope is consistent with changes in estimated variability directly reflecting corresponding changes in true variability; but the responses show a consistent and constant bias, underestimating variability with all the points below the perfect calibration line. The flat line in the figure might characterise the performance of an individual who is 'variability-blind', but who is 'just guessing' and chooses, for whatever reason, to make a high-variability estimation. We discuss some possible reasons in our model below. To clearly assure that a measure of response is meaningful, one must determine that at least *some* sensitivity to variability differences is exhibited.

These different types of measures also introduce a note of caution concerning findings of over- versus under-estimation of variability. Studies reporting bias often report participants' estimated variability for only a single level of true variability, but as can be seen from the arrows in the center panel of Figure 2, at different levels of true variability different patterns of over- and underestimation of variability could have been possible, whenever sensitivity is not perfect. Thus, the presence of overconfidence (underestimating variability) in some contexts may reflect the combination of the task and the chosen level of true variability presented to the judger, rather than a general tendency towards overconfidence in mean estimation. In addition, any research that only employs measures of precision offers no hint on directionality, and thus does not speak to this important issue.

In order to provide an underlying theoretical framework for understanding the biases in estimating variability, in the following section we unpack the different cognitive mechanisms that are likely responsible for generating these estimates, and how other theories of perception, cognition (memory and bias), and response may be responsible for moderating the variability estimates. This can be considered the foundation of a MOVE.

5. The model of variability estimation

Consider a simplified representation of the multiple instances that might be encountered in any paradigm, or in real world experience, whose variability is to be later inferred (after all instances have been encountered). Those instances are all observations arrayed along some continuously distributed quantity (ratio *or* interval scale) but may appear in either analogue spatial form (e.g. hurricane trend lines) or numeric form (e.g. reporting of stock trend numbers). Also, the distribution of instances may have any level of true variability. How accurately people can encode, retain and recall (i.e. judge) the actual levels of variability in these cases depends on how well they can discriminate a low from a high variability set of instances, as shown by the thick solid line sensitivity functions in Figure 2.

In our model of this judgement/discrimination process, adopting a conventional learning/memory representation, we distinguish between features of *encoding, retention, bias and retrieval* (judgement measurement), identify elements within each that affect the quality of processing therein; and how these may then be expressed in the three outcome *measures of performance* depicted in Figure 2: sensitivity, bias and precision. The influence of some of these factors is inferred from the literature presented above. Other variables can be posited to influence based upon research from other related domains. We describe these influences as follows.

5.1. Encoding

In most typical situations, a mental model of variability must depend upon being exposed to multiple instances. There are several features that characterise the experience of instance encoding including:

1. Familiarity of the material. Numbers may be more familiar (and hence better encoded) than points distributed along a line. Also estimating the current state of each instance is probably a more familiar (and easier) task than predicting the future state, as in trajectory extrapolation (e.g. hurricane paths). Thus, given the influence of familiarity on memory encoding, we posit that greater sensitivity, and hence greater precision will result from more familiar instances (Reder et al. 2007), and we have found this to be true, with digital material showing better calibration than spatial material (Spahr et al. 2018; Spahr 2019).
2. Dimensionality of instances. Variability may sometimes occur along one dimension alone. However, when instances are multidimensional (for example projecting the location, time and severity of hurricane landfall), we postulate that the greater number of dimensions, the more the encoding and representation of each dimension will be degraded (lower sensitivity) by dividing attention. This reflects basic findings of multi-dimensional stimulus processing (e.g. Garner 1974).
3. Correlation of dimensions. For multidimensional instances, encoding will differ depending on whether or not the dimensions are uncorrelated across instances or are

correlated (e.g. more intense hurricanes tend to move slower). Correlated dimensions tend to be better encoded (Garner 1974), and hence might be expected to generate a more reliable judgement of variability.

4. Attention paid to instances. Greater attention could promote better (more enduring) encoding of variability information (Craik and Lockhart 1972), but there are several circumstances that can influence the amount of resources allocated to variability of instances, including:
   - Instructions to focus attention on such variability (Herdener et al. 2018).
   - Concurrent distracting tasks, such as either a loading task or a task, such as prediction of the mean instance, that may divert attention away from encoding the variability across instances (Herdener et al. 2018).
   - Financial incentives for accurate judgement of variability (Spahr 2019).
   - Visualisation attentional focus techniques, such as a cone of uncertainty or 'spaghetti plot' (Broad et al. 2007; Cox, House, and Lindell 2013; Padilla, Ruginski, and Creem-Regehr 2017; Pugh et al. 2018; Ruginski et al. 2016; Boone, Gunalp, and Hegarty 2018).

   Regarding these attentional issues, our research to this point has shown that directing more attention to variability does *not* consistently improve the quality of later judgement (Herdener et al. 2018), as if this cognitive task is somewhat 'data limited' rather than resource-limited (Norman and Bobrow 1975). However, there is some evidence that both financial incentives for calibration and visualisation may do so (Spahr et al. 2018; Spahr 2019; Herdener et al. 2019a)

   One possibility is that the role of attention is not that a focus on variability is directly capturing an enhanced sense of the variability across instances. Rather in line with Logan's Instance Theory (1988), attention to information engages obligatory encoding of the instances that occur and attending to a stimulus also invokes the retrieval of related previous instances from memory. How attention is being directed might change the types of instances being retrieved, and hence the resulting understanding of variability.

5. Temporal nature of encoding presentation. Encounters with multiple instances may be concurrent, or sequential with each of instance presented one at a time. In sequential presentations, the time span might range from several instances per second in an animation, to an extreme, outside the lab, in judging the variability of severity of hurricanes, in which a person may be exposed to perhaps three trajectories per month during hurricane season. The temporal nature can be defined jointly by the *inter-instance*-interval and the *total instance* exposure time. These two measures will not be perfectly correlated across differences in the number of instances encountered. As we see below, temporal nature will affect the memory system that is used to retrieve and hence judge variability, and through this, should affect the quality of that response with longer time precluding the use of working memory, and thereby producing lower resolution encoding.

6. Size (variance) of the distribution. This feature is a direct assessment of the standard deviation or range of instance. For distributions of spatial stimuli, it can be represented in terms of the breadth of visual angle, or the proportion of a display screen occupied by the set of instances. For temporal uncertainty it can also be defined by standard deviations of time (Jobidon, Rousseau and Breton 2005). Variability thus can be subject to threshold limits, so that small variability is not itself well resolved,

and hence differences between two low variability distributions cannot easily be discriminated (Spahr et al. 2018). Alternatively, it may be that variability follows a Weber's Law psychophysical function: the assessments of differences in variability scales more or less linearly with the absolute amount of such variability.

7. Parameters of the distribution of instances: this has two important subcomponents. First, this includes the raw amount of variability as described above; Second, this importantly, includes the shape of the distribution itself (Beach and Scopp 1968; Rinne and Mazzoco 2013). For example, encoding variability may change depending on whether a distribution is rectangularly or normally distribute, or more platykurtic or leptokurtic. One can postulate that the normal distribution, by presenting relatively more instances packed around the mean, would lead to smaller estimations of variability than a rectangular one (Rinne and Mazzoco 2013).
8. Number of instances. This could be considered analogous to sample size in an experiment. If the human is a 'perfect integrator', more samples should lead to closer calibration. However, if the human's judgement of variability is only based on a running memory span of the last few instances (Herdener et al. 2019a; Kareev, Arnon, and Horwitz-Zeliger 2002), then the number should play little role, above the first few instances encountered (e.g. 6–10 s worth).
9. Order of presentation: for sequential instances, this would include for example whether the sequence is random over time, or is systematic (e.g. increasing from low to high variability, Lathrop 1967); or the extent to which extreme values are placed at the beginning or end of a sequence of instances. In these cases, heuristics of anchoring (primacy) or recency (Hogarth and Einhorn 1992; Entin and Serfaty 1997; Wickens et al. 2010) may influence memory recall, as we discuss below.

5.2. Cognition: working memory and retention interval for multiple instance exposure

It is well established from memory research that, in the absence of rehearsal, accuracy of recall declines with increasing retention intervals, generally following an exponential decay curve (Ebbinghaus 1913).

The temporal nature of instance presentation (see above) is relevant here, as it was with encoding. If the presentation is simultaneous, then this interval is unambiguous: it is the delay between exposure termination and variability retention assessment. But for sequential presentations it becomes more complex as shown in Figure 3, which depicts the time line of five instances (*) and a retention test (*R*) and illustrates three different ways of defining the retention interval.

In the top line, the retention interval 'clock' is started from the first instance. In the second line, the clock is started from the halfway point (when half the instances are presented), and in the bottom line, the clock is started from the final instance. Which of these is used matters if there are variations in instance presentation rate (See 'Encoding' above).

We would argue that the key feature of retention interval that can affect the quality of recall is the number of instances that remain in working memory or iconic memory at the time that judgement is made. There will be more of these to the extent that the retention interval after the last instance is short, and/or that the presentation rate is rapid. But if the

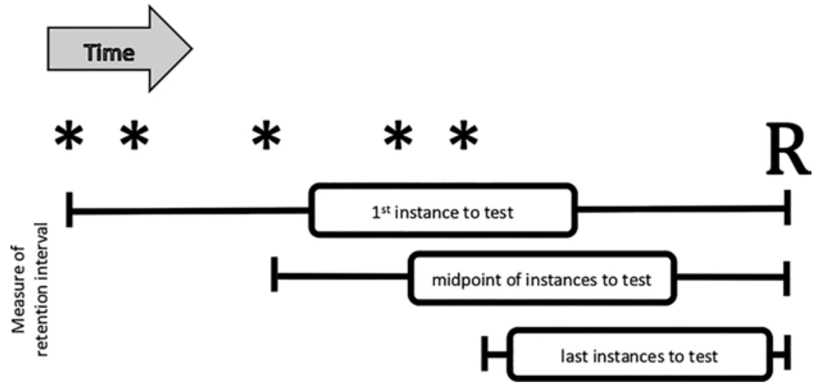

Figure 3. Alternative definitions of the retention interval. The stars represent a sequence of instances. 'R' represents the time of retention measurement. The three lines represent different ways of timing the retention interval, as discussed in the text below.

retention interval since the last instance is longer than, perhaps 10 s (an upper bound for unrehearsed working memory duration) then further delays, while causing a gradual loss of judgement precision, will also lead to a more gradual slope of loss, as reflecting the longer forgetting time constant of long term memory (Card, Moran, and Newell 1986).

We infer that this differential influence of working memory can account for the better calibration of variability estimation observed in the Spahr et al. (2018) digit memory task, in which three to four digits remain in working memory at judgement time, than in the tasks used by Herdener et al. (2016, 2018, 2019a) where the delay was longer, in the order of minutes.

5.3. Cognition: biases during retention

An ideal estimate of variability would accumulate all deviations of each instance from the mean, and mentally average these after the last instance is encountered. The precision of this estimation will depend on the fidelity of memory.

However, there are many biases that could impact the estimate. One prominent example is *anchoring* (and failure to subsequently adjust enough; Tversky and Kahneman 1974; Hogarth and Einhorn 1992; Entin and Serfaty 1997; Adelman et al. 1996; Wickens et al. 2010). Here we can postulate at least three kinds of anchors:

Initial anchor: the variability between the first few instances plants an initial estimate of variability. For example, if instances 1 and 2 are widely separated, the initial anchor will create a mental model of high variability. This is what, in memory research, is classically called a *primacy effect*. With respect to perceiving variability, Witt (2019) showed that judgements of the variability in the orientations of a set of lines was more influenced by the first lines to be shown than by the later lines.

Final anchor. This is the reverse of the initial anchor and will predict excessive influence of the variability across the final three to four instances (and more instances if the inter-instance-interval is shorter). This is the *recency effect*, (Hogarth and Einhorn 1992; Entin and Serfaty 1997; Wickens et al. 2010) and will probably diminish as the retention interval is longer. With respect to perceiving the mean angle of a set of lines, Witt (2019) showed that the last lines to be presented had more influence on mean judgements, so a primacy effect was found for judging orientation variability and a recency effect was found for judging mean angle.

Task (mean)-driven anchor. A task that either explicitly or implicitly encourages estimating the mean of sequential samples (for examples see, Herdener et al. 2016, 2018, 2019a; Spahr et al. 2018) plants an anchor towards the middle of the distribution (and a potentially stronger anchor for people who have greater accuracy in estimating the mean). Hence adjustment on the basis of subsequent instances will be pulled towards this mean and will result in an under-estimation of variability. Influences related to instance sequencing in variability estimation have been shown by Lathrop (1967). Some evidence to support this is provided by the findings that those individuals who do better estimating the mean, actually so worse in estimating variability (Herdener et al. 2016).

5.4. Retention assessment: the response methodology

Different paradigms vary in how they examine the implications of variability judgement biases for subsequent decisions. For example, one can examine the likelihood of evacuating regions around the cone of uncertainty in a simulated hurricane approach (Ruginski et al. 2016; Boone, Gunalp, and Hegarty 2018). However, to more directly examine the calibration of judgement (rather than the downstream implications of mis-calibration for decision making), we have employed two different methodologies as shown in Figure 4.

On the left of the figure, the participant sees and encodes several instances of the termination location of a trajectory or other quantitative variable. The mental model of their experience of variability across those instances (shown within the cloud) is then assessed. This mental model may overestimate (top cloud) or underestimate (bottom cloud) the true amount of variability. Alternatively, of course, it may be accurately calibrated. The mental model is then assessed by one of two judgement response techniques: (1) Asking participants to *estimate* the percentage within an interval or circle of fixed size. The circle in the figure would contain 60% of the trajectory endpoints, but estimates tend to be around 80% as shown in the bottom number, indicating an underestimation of variability. (2) Asking them to *adjust* a one-dimensional bracket or two-dimensional circle (shown here) to encompass a specified percentage of the instances encountered (in this example, 60%). We may think of this as producing a subjective confidence interval, a technique used in several of the paradigms discussed above. With either the estimation or adjustment response technique to elicit the judgement, the pattern of *underestimating* variability is that produced at the bottom right, and the pattern of *overestimation* is shown at the top right.

Hansson, Juslin, and Winman (2008) observes more consistent findings of variability underestimation with the technique of confidence interval adjustment, than with other techniques. Rinne and Mazzoco (2013) find that people's mental model of a distribution is less accurate with larger confidence intervals (i.e. 90 or 95%), than with smaller ones (i.e. 75%, 40%). Teigen and Jørgensen (2005) observe that greater overconfidence is produced with these larger confidence intervals. Neither of these studies, however, directly assessed variability estimation of multiple instances, whose presentation parameters were controlled.

The differential effect of these two methodologies on precision, sensitivity and bias is not well understood. In our own research, we tend to find that the methods of estimation

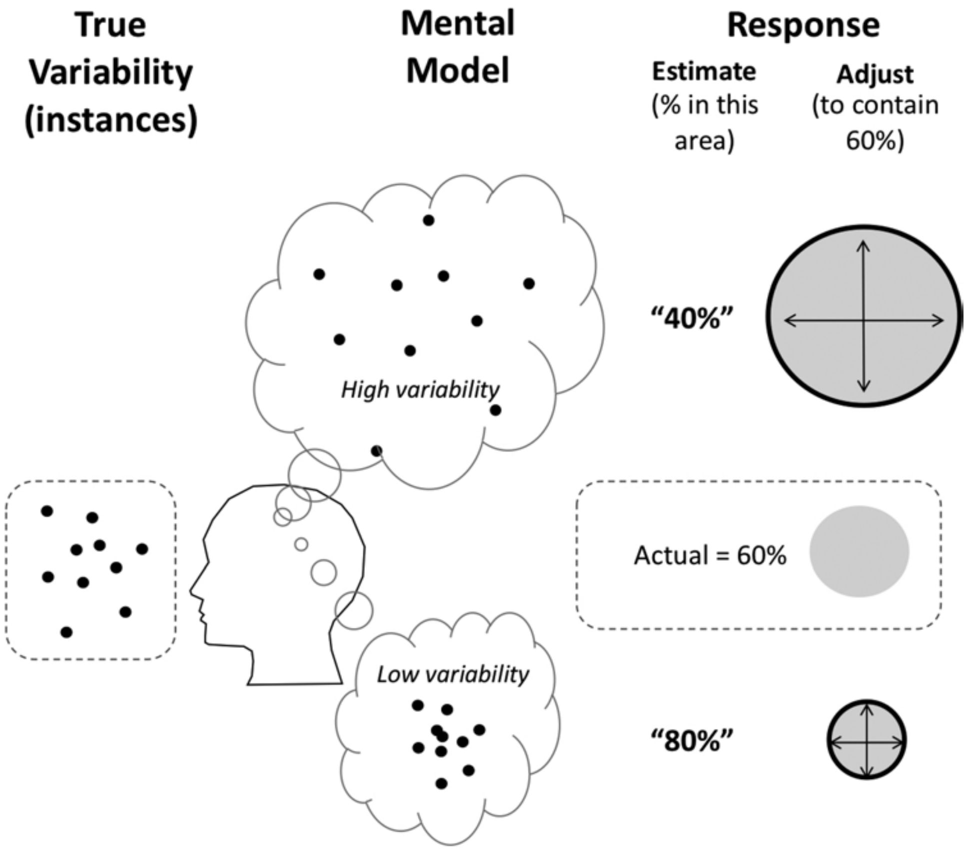

Figure 4. Response methodologies for assessing variability judgements. Explanation of elements is contained in the text.

produce a greater bias towards underestimation of variability, than the method of adjustment (Spahr 2019), and that adjustment is the more sensitive measure to reflect variability differences (Herdener et al. 2019a). However, it may well be the case that the method of response adjustment is affected by the tools used to generate the response. For example, a tool that is difficult to create larger intervals will produce a bias towards smaller (under-estimated) variability. Furthermore, it is possible that the differences in results between the two techniques may represent the more abstract requirement for estimation to require a numerical response, whereas adjustment can require the more natural spatial-manual response, directly compatible with the spatial stimuli with which it has been employed.

5.4.1. Individual differences

As discussed by Pallier et al. (2002), individual differences in cognitive abilities may account for some differences in overconfidence; hence, given the linkage described above between overconfidence in precision, and underestimation of variability we might infer it to underlie differences in variability estimation.

Working memory. One contributing factor may be individual differences in working memory, a cognitive ability well-documented to impact a variety of other processes. For example, working memory correlates with Level 3 Situational Awareness, responsible for projection, or predicting a future state based on current assessment and understanding (Gutzwiller and Clegg 2013; Sulistyawati, Wickens, and Chui 2011). Further, Hansson, Juslin, and Winman (2008) observed that overconfidence in the accuracy of knowledge judgements was negatively correlated with WM capacity, such that those with higher WM demonstrated better calibration. Kareev, Arnon, and Horwitz-Zeliger (2002) correspondingly found that variability estimation quality is correlated with the capacity of working memory. Herdener et al. (2018) found that variability estimation could be best modelled with a very limited working memory system. To accurately assess variability, particularly over time, we argue that one must encode the differences between *multiple* instances and manipulate these instances, suggesting the use of working memory.

Numeracy. Past research commonly applies the understanding of numbers and computation, or numeracy, in learning theory and comprehension of medical trends. Individual differences in numeracy may also impact variability judgements, specifically those numeric in nature. Rinne and Mazzoco (2013) examined the relationship between numeracy and estimates of uncertain numeric intervals; their results suggest evidence that high numeracy helped make accurate estimates, but this effect was most pronounced when inferring a normal distribution.

Expertise. Past research has often examined professionals deemed 'expert' in prediction and assessment of both discrete events and ongoing trends, such as highly trained financial analysts and weather forecasters. Individual differences in expertise (i.e. the expert versus the novice) may contribute to accuracy of such predictive judgements and the role of overconfidence. High skill and ability may be reflected in estimating or predicting the mean; however, this may not translate to better predicting variability or may even hurt it, as we have seen from the poor estimates of even highly trained financial forecasters (Silver 2012; Einhorn and Hogarth 1978). Further, individual differences between experts in different domains provide a compelling case: Tyszka and Zielonka (2002) demonstrated that while both highly trained financial analysts and weather forecasters were overconfident, financial analysts showed significantly higher overconfidence in their predictions. These overconfidence findings were ascribed to learning from experience (i.e. the uncertain nature of the domain of prediction), cognitive heuristics and experts' self-motivation. In sum, individual differences in expertise could account for trends in both performance and overconfidence. However, the extent to which increased training (i.e. the production of expertise) increases variability judgement precision, sensitivity, or bias, cannot be ascertained from the currently available research.

5.5. Features of a MOVE

Based upon our review of the literature in Section 2, and our identification of factors found, or likely to influence the judgement of variability described in Section 3, we have tried to extract from these, the foundations of a model of these influences. In Table 1, the three main categories of influential variables are shown (perception, cognition, response), with each category broken down by specific experimental manipulations.

Notes. '+' indicates increasing, '–' indicates decreasing. '0' indicates no effect. Double symbols indicate greater confidence that the effect is true and/or larger effect. Blank cells indicate the information is unknown. Cells with two opposing symbols are cases for which

Table 1. Different factors found to influence variability estimation.

Factors	Effect on sensitivity	UEV to OEV bias	Reference
Stimulus (perception)			
Familiarity	+		Spahr
# Dimensions	–		
Redundant dimensions	+		
Attention paid	+	0	Herdener 18a
Distribution: size	++	––	Spahr
Shape: rectangle to normal		–	Rinne, Beach and Scopp
# Instances	+	+	
Multiple exposure time	––		Witt
Cognition (memory and bias)			
Retention interval	––		Spahr
Mean anchor		–	
Response			
Estimation to adjustment (CI)	–	0 –	Spahr, Hansson
Increasing size of CI	–	–	Teigen
Numeracy	+		Hansson
Working memory capacity	++	–	Hansson, Kareev

Notes. '+' indicates increasing, '–' indicates decreasing. '0' indicates no effect. Double symbols indicate greater confidence that the effect is true and/or larger effect. Blank cells indicate the information is unknown. Cells with two symbols are cases for which different studies produced opposing conclusions. Last column indicates first name of author (in reference) of paper that is basis of inference.

different studies produced opposing conclusions. Last column indicates first name of author (in reference) of paper that is basis of inference.

The strength of effect of increasing the variable in question on sensitivity to variability differences (a necessary prerequisite for precision) is shown in the next column. [+] means it will increase precision and sensitivity, [–] means it will decrease it. We have estimated this magnitude jointly from the consistency of findings and the strength of effect when reported.

The inferred influence on bias is shown in the next column. '+' means that the variable in question is likely to either reduce an underestimation bias or enhance an overestimation bias. A '–' means the converse.

The coarse-grained estimated effect sizes shown in Table 1 could, as a next step, be combined to create a linear additive model of the form:

$$\text{Sensitivity} = aA + bB + cC$$
$$\text{Bias} = uU + vV + wW....,$$

where each letter is the source of influence on sensitivity or bias (a negative influence on sensitivity would be assumed to cascade to a negative effect on precision). The capital letter represents how a given paradigm 'scores' on the influence factor in question, and the sign indicates whether the influence is positive or negative (e.g. improves or degrades precision).

At this point, following the guidance of Dawes (1979; Dawes and Corrigan 1974), we can assume very simple equal weightings on all factors (all lower-case parameters = 1.0), or, at most, a coarse weighting such as reflected in Table 1.

We can also assume, for simplicity, linear functions for all variables as a starting point. The one exception here is the negative weighting on retention interval which can be estimated as an exponential decay function, as discussed above, reflecting the progressively longer time constants for iconic, working and long term memory, respectively (Card, Moran, and Newell 1986). The authors are anticipating that systematic experiments will be conducted to manipulate these factors independently.

6. Conclusions

In closing, we might ask if there is a biologically adaptive reason, above and beyond the information processing challenges presented in our model, why people might generally underestimate variability. It seems reasonable to infer that, because so many tasks in life depend upon averaging, 'typifying' or 'generalising' to form a prototype, and hence rapidly eliminating a working memory load, for which variability of multiple instances is simply a nuisance, that humans, and perhaps other species may be 'tuned' to ignore variability.

In conclusion, we have argued that there are many safety critical domains, from process control to hurricane evacuation in which correct prediction of continuous trends has a critical impact on the discrete action taken (and must be initiated early if there is considerable time required to implement it). We have also provided evidence that people do not do this task very well, particularly when prediction of slowly changing trends involves accounting for future uncertainty (variability) of such trends. In particular, there is evidence of a tendency to under-estimate that variability which, like a *miss* in signal detection, can have consequences that are more dire than its counterpart of overestimation (or a false alarm in signal detection).

In contrast to other ergonomically relevant cognitive processes, such as vigilance or discrete decision under uncertainty, the area of continuous trend variability estimation has received very little empirical investigation and, to our knowledge has never been the target of empirically based quantitative modelling. The 'model' we propose here is tentative and preliminary, but we believe provides the foundation for a program of research that will have considerable payoffs for ergonomic domains, where calibrated prediction can be the cornerstone for safe actions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Dr Jeff Morrison and the Office of Naval Research under grant No. 00014-17-1-2825 and by a grant from the National Science Foundation (BCS-1632222) to JKW.

References

Adelman, Leonard, Terry Bresnick, Paul K. Black, F. Freeman Marvin, and Steven G. Sak. 1996. "Research with Patriot Air Defense Officers: Examining Information Order Effects." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38 (2): 250–261. doi:10.1177/001872089606380206.

Beach, Lee Roy, and Thomas S. Scopp. 1968. "Intuitive Statistical Inferences about Variances." *Organizational Behavior and Human Performance* 3 (2): 109–123. doi:10.1016/0030-5073(68)90001-9.

Boone, Alexander P., Peri Gunalp, and Mary Hegarty. 2018. "Explicit versus Actionable Knowledge: The Influence of Explaining Graphical Conventions on Interpretation of Hurricane Forecast Visualizations." *Journal of Experimental Psychology: Applied* 24 (3): 275–295. doi:10.1037/xap0000166.

Broad, Kenneth, Anthony Leiserowitz, Jessica Weinkle, and Marissa Steketee. 2007. "Misinterpretations of the "Cone of Uncertainty" in Florida during the 2004 Hurricane Season." *Bulletin of the American Meteorological Society* 88 (5): 651–668. doi:10.1175/BAMS-88-5-651.

Card, Stuart, Thomas Moran, P., and Allen Newell. 1986. "The Model Human Processor: An Engineering Model of Human Performance." In *Handbook of Perception and Performance,* edited by Boff, Kenneth R., Lloyd Kaufman, and James Thomas, (2) 45. New York: John Wiley & Sons.

Cox, Jonathan, Donald House, and Michael Lindell. 2013. "Visualizing Uncertainty in Predicted Hurricane Tracks." *International Journal for Uncertainty Quantification* 3 (2): 143–156. doi:10.1615/Int.J.UncertaintyQuantification.2012003966.

Craik, Fergus I.M., and Robert Lockhart. 1972. "Levels of Processing: A Framework for Memory Research." *Journal of Verbal Learning and Verbal Behavior* 11 (6): 671–684. doi:10.1016/S0022-5371(72)80001-X.

Dawes, Robyn M. 1979. "The Robust Beauty of Improper Linear Models in Decision Making." *American Psychologist* 34 (7): 571–582. doi:10.1037/0003-066X.34.7.571.

Dawes, Robyn M., and Bernard Corrigan. 1974. "Linear Models in Decision Making." *Psychological Bulletin* 81 (2): 95–106. doi:10.1037/h0037613.

DeLucia, Patricia R. 2015. "The Perception of Collisions." In *The Cambridge Handbook of Applied Perception Research*. Vol. 1, edited by Robert R. Hoffman, Peter A. Hancock, Mark W. Scerbo, Raja Parasuraman, and James L. Szalma, 568–591. Cambridge, UK: Cambridge University Press.

Ebbinghaus, Hermann. 2013. "Memory: A Contribution to Experimental Psychology." *Annals of Neurosciences* 20 (4): 155–156.

Ebbinghaus, Hermann. 1913. *Memory: A Contribution to Experimental Psychology.* Translation by H. A. Ruger and C. E. Bussenius. New York: Teachers College Press.

Einhorn, Hillel J., and Robin M. Hogarth. 1978. "Confidence in Judgment: Persistence of the Illusion of Validity." *Psychological Review* 85 (5): 395–416. doi:10.1037/0033-295X.85.5.395.

Engström, Johan, Jonas Bärgman, Daniel Nilsson, Bobbie Seppelt, Gustav Markkula, Giulio Piccinini, and Trent Victor. 2018. "Great Expectations: A Predictive Processing account of Automobile Driving." *Theoretical Issues in Ergonomics Science* 19 (2): 156–194. doi:10.1080/1463922X.2017.1306148.

Entin, Elliot E., and Daniel Serfaty. 1997. "Sequential Revision of Belief: An Application to Complex Decision Making Situations." *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 27 (3): 289–301. doi:10.1109/3468.568738.

Fischoff, Baruch, Paul Slovic, and Sarah Lichtenstein. 1978. "Fault Trees: Sensitivity of Estimated Failure Probabilities to Problem Representation." *Journal of Experimental Psychology: Human Perception and Performance* 4 (2): 330–344. doi:10.1037//0096-1523.4.2.330.

Garner, Wendell R. 1974. *The Processing of Information and Structure*. Hillsdale N.J.: Erlbaum Associates.

Gutzwiller, Robert S., and Benjamin A. Clegg. 2013. "The Role of Working Memory in Levels of Situation Awareness." *Journal of Cognitive Engineering and Decision Making* 7 (2): 141–154. doi:10.1177/1555343412451749.

Hamilos, Christopher A., and Gordon F. Pitz. 1977. "The Encoding and Recognition of Probabilistic Information in a Decision Task." *Organizational Behavior and Human Performance* 20 (2): 184–202. doi:10.1016/0030-5073(77)90002-2.

Hansson, Patrik, Peter Juslin, and Anders Winman. 2008. "The Role of Short-Term Memory Capacity and Task Experience for Overconfidence in Judgment Under Uncertainty." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 34 (5): 1027–1042. doi:10.1037/a0012638.

Henrion, Max, and Baruch Fischhoff. 1986. "Assessing Uncertainty in Physical Constants." *American Journal of Physics* 54 (9): 791–798. doi:10.1119/1.14447.

Herdener, Nathan, Christopher D. Wickens, Benjamin Clegg, and C. A. P. Smith. 2018. "Attention Does Not Improve Impaired Understanding of Variability in Spatial Prediction." In. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62 (1): 232–236. doi:10.1177/1541931218621054.

Herdener, Nathan, Benjamin A. Clegg, Christopher D. Wickens, and C. A. P. Smith. 2019a. "Anchoring and Adjustment in Uncertain Spatial Trajectory Prediction." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 61 (2): 255–272. doi:10.1177/0018720818800591.

Herdener, Nathan, Benjamin A., Clegg, Christopher D. Wickens, and C. A. P. Smith. 2019b. "Can You See It? Perceived Variance in Scatterplot Visualization." In *Proceedings of the 2019 Annual Conference of the Human Factors and Ergonomics Society*.

Herdener, Nathan, Christopher D. Wickens, Benjamin A. Clegg, and C. A. P. Smith. 2016. "Overconfidence in Projecting Uncertain Spatial Trajectories." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58 (6): 899–914. doi:10.1177/0018720816645259.

Hogarth, Robin M., and Hillel J. Einhorn. 1992. "Order Effects in Belief Updating: The Belief-Adjustment Model." *Cognitive Psychology* 24 (1): 1–55. doi:10.1016/0010-0285(92)90002-J.

Jobidon, Marie-Eve, Robert Rousseau, and Richard Breton. 2005. "The Effect of Temporal Information on the Control of a Dynamic Task." *Theoretical Issues in Ergonomics Science* 6 (1): 49–62. doi:10.1080/14639220512331311562.

Juslin, Peter, Anders Winman, and Patrik Hansson. 2007. "The Naïve Intuitive Statistician: A Naïve Sampling Model of Intuitive Confidence Intervals." *Psychological Review* 114 (3): 678–703. doi:10.1037/0033-295X.114.3.678.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kareev, Yaakov, Sharon Arnon, and Reut Horwitz-Zeliger. 2002. "On the Misperception of Variability." *Journal of Experimental Psychology: General* 131 (2): 287–297. doi:10.1037/0096-3445.131.2.287.

Lathrop, Richard G. 1967. "Perceived Variability." *Journal of Experimental Psychology* 73 (4, Pt.1): 498–502. doi:10.1037/h0024344.

Levin, Irwin P., John R. Ims, John C. Simpson, and Kyung Ja Kim. 1977. "The Processing of Deviant Information in Prediction and Evaluation." *Memory & Cognition* 5 (6): 679–684. doi:10.3758/BF03197415.

Logan, Gordon D. 1988. "Toward an Instance Theory of Automatization." *Psychological Review* 95 (4): 492–527. doi:10.1037/0033-295X.95.4.492.

Lovie, Patricia, and A.D. Lovie. 1976. "Teaching Intuitive Statistics 1: Estimating Means and Variances." *International Journal of Mathematical Education in Science and Technology* 7 (1): 29–39. doi:10.1080/0020739760070104.

Lovie, Patricia. 1978. "Teaching Intuitive Statistics 2: Aiding Estimates of Standard Deviation." *International Journal of Mathematical Education in Science and Technology* 9 (2): 213–219. doi:10.1080/0020739780090210.

McCarthy, Gerard D., Ivan D. Haigh, Joël J.-M. Hirschi, Jeremy P. Grist, and David A. Smeed. 2015. "Ocean Impact on Decadal Atlantic Climate Variability Revealed by Sea-Level Observations." *Nature* 521 (7553): 508. doi:. doi:10.1038/nature14491.

Moore, Don A., and Paul J. Healy. 2008. "The Trouble with Overconfidence." *Psychological Review* 115 (2): 502–517. doi:10.1037/0033-295X.115.2.502.

Muhs, Kevin, Waldemar Karwowski, and Dave Kern. 2018. "Temporal Variability in Human Performance: A Systematic Literature Review." *International Journal of Industrial Ergonomics* 64: 31–50. doi:10.1016/j.ergon.2017.10.002.

Norman, Donald A., and Daniel G. Bobrow. 1975. "On Data-Limited and Resource-Limited Processes." *Cognitive Psychology* 7 (1): 44–64. doi:10.1016/0010-0285(75)90004-3.

Obrecht, Natalie A., Gretchen B. Chapman, and R. Gelman. 2007. "Intuitive T-Tests: Lay Use of Statistical Information." *Psychonomic Bulletin & Review* 14 (6): 1147–1152. doi:10.3758/BF03193104.

Olsson, Henrik. 2014. "Measuring Overconfidence: Methodological Problems and Statistical Artifacts." *Journal of Business Research* 67 (8): 1766–1770. doi:10.1016/j.jbusres.2014.03.002.

Padilla, Lace M., Ian T. Ruginski, and Sarah H. Creem-Regehr. 2017. "Effects of Ensemble and Summary Displays on Interpretations of Geospatial Uncertainty Data." *Cognitive Research: Principles and Implications* 2 (1): 40. doi:10.1186/s41235-017-0076-1.

Pallier, Gerry, Rebecca Wilkinson, Vanessa Danthiir, Sabina Kleitman, Goran Knezevic, Lazar Stankov, and Richard D. Roberts. 2002. "The Role of Individual Differences in the Accuracy of Confidence Judgments." *The Journal of General Psychology* 129 (3): 257–299. doi:10.1080/00221300209602099.

Pollard, P. 1984. "Intuitive Judgments of Proportions, Means, and Variances: A Review." *Current Psychological Research & Reviews* 3 (1): 5–18. doi:10.1007/BF02686528.

Peterson, Cameron R., and Lee Roy Beach. 1967. "Man as an Intuitive Statistician." *Psychological Bulletin* 68 (1): 29–46. doi:10.1037/h0024722.

Pitz, Gordon F. 1980. "The Very Guide of Life: The Use of Probabilistic Information for Making Decisions." In *Cognitive Processes in Choice and Decision Behavior,* edited by Thomas S. Wallsten, Hillel J. Einhorn, Ebbe B. Ebbesen, Vladimir J. Konecni, and Ruth B. Corbin, 77–94. Hillsdale N.J.: Erlbaum & Associates.

Pugh, Ashley J., Christopher D. Wickens, Nathan Herdener, Benjamin A. Clegg, and C. Smith. 2018. "Effect of Visualization Training on Uncertain Spatial Trajectory Predictions." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60 (3): 324–339.

Ray, Deepak K., James S. Gerber, Graham K. MacDonald, and Paul C. West. 2015. "Climate Variation Explains a Third of Global Crop Yield." *Nature Communications* 6 (1). 5989. doi:10.1038/ncomms6989.

Reder, Lynne M., Paynter, Christopher, and Rachel A. Diana, Jiquan Ngiam, Daniel Dickison. 2007. "Experience Is a Double-Edged Sword: A Computational Model of the Encoding/Retrieval Trade-Off with Familiarity." *Psychology of Learning and Motivation* 48: 271–312.

Rinne, Luke F., and Michèle Mazzoco. 2013. "Inferring Uncertainty from Interval Estimates: Effects of Alpha Level and Numeracy." *Judgment and Decision Making* 8 (3): 330–344.

Ruginski, Ian T., Alexander P. Boone, Lace M. Padilla, Le Liu, Nahal Heydari, Heidi S. Kramer, Mary Hegarty, William B. Thompson, Donald H. House, and Sarah H. Creem-Regehr. 2016. "Non-Expert Interpretations of Hurricane Forecast Uncertainty Visualizations." *Spatial Cognition & Computation* 16 (2): 154–172. doi:10.1080/13875868.2015.1137577.

Silver, Nate. 2012. *The Signal and the Noise: Why so Many Predictions Fail–but Some Don't*. New York: Penguin Press.

Soll, Jack B., and Joshua Klayman. 2004. "Overconfidence in Interval Estimates." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30 (2): 299–314. doi:10.1037/0278-7393.30.2.299.

Suss, Joel, and Paul Ward. 2015. *Predicting the Future in Perceptual-Motor Domains." in the Cambridge Handbook of Applied Perception Research,* edited by James L. Szalma, Mark W. Scerbo, Raja Parasuraman, Peter A. Hancock, Robert R. Hoffman, 951–976. New York: Cambridge University Press.

Spahr, Kimberly S. 2019. "*Estimating Variability across Numeric and Spatial Information.*" Unpublished master's thesis, Colorado State University.

Spahr, Kimberly S., Christopher D. Wickens, Benjamin A. Clegg, C. A. P. Smith, and Adam S. Williams. 2018. "Calibrating Uncertainty: Commonalities in the Estimation of Numeric Variability versus Spatial Prediction." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62 (1): 757–761. doi:10.1177/1541931218621172.

Strobhar, David A. 2014. *Human Factors in Process Plant Operation*. New York: Momentum Press.

Sulistyawati, Ketut, Christopher D. Wickens, and Yoon Ping Chui. 2011. "Prediction in Situation Awareness: Confidence Bias and Underlying Cognitive Abilities." *The International Journal of Aviation Psychology* 21 (2): 153–174. doi:10.1080/10508414.2011.556492.

Teigen, Karl Halvor, and Magne JØrgensen. 2005. "When 90% Confidence Intervals Are 50% Certain: On the Credibility of Credible Intervals." *Applied Cognitive Psychology* 19 (4): 455–475. doi:10.1002/acp.1085.

Tetlock, Philip, E., 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton N.J.: Princeton University Press.

Tversky, Amos, and Daniel Kahneman. 1974. “Judgment under Uncertainty: Heuristics and Biases.” *Science* 185 (4157): 1124–1131. doi:10.1126/science.185.4157.1124.
Tyszka, Tadeusz, and Piotr Zielonka. 2002. “Expert Judgments: Financial Analysts versus Weather Forecasters.” *Journal of Psychology and Financial Markets* 3 (3): 152–160. doi:10.1207/S15327760JPFM0303_3.
Whitney, David, and Allison Yamanashi Leib. 2018. “Ensemble Perception.” *Annual Review of Psychology* 69 (1): 105–129. doi:10.1146/annurev-psych-010416-044232.
Wickens, Christopher D., Shaw L. Ketels, Alice F. Healy, Carolyn J. Buck-Gengler, and Lyle E. Bourne. 2010. “The Anchoring Heuristic in Intelligence Integration: A Bias in Need of De-Biasing.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54 (27): 2324–2328. doi:10.1177/154193121005402722.
Wickens, Christopher. D., Adam Williams, Cap Smith, and Benjamin A. Clegg. 2019. “Nautical Collision Avoidance in Uncertainty: The Cognitive Challenges of Balancing Safety, Efficiency and Procedures.” *Human Factors*.
Wickens, Thomas D. 2002. *Signal Detection Theory*. San Francisco, CA: Freeman.
Witt, Jessica K. 2019. “The Perceptual Experience of Variability in Line Orientation Is Greatly Exaggerated.” *Journal of Experimental Psychology: Human Perception and Performance* 45 (8): 1083–1103.