

Environmental risk perception from visual cues: the psychophysics of tornado risk perception

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Environ. Res. Lett. 10 124009

(<http://iopscience.iop.org/1748-9326/10/12/124009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 71.112.208.157

This content was downloaded on 08/12/2015 at 13:55

Please note that [terms and conditions apply](#).

Environmental Research Letters



LETTER

Environmental risk perception from visual cues: the psychophysics of tornado risk perception

OPEN ACCESS

RECEIVED

25 April 2015

REVISED

25 September 2015

ACCEPTED FOR PUBLICATION

30 September 2015

PUBLISHED

8 December 2015

Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Barry Dewitt^{1,3}, Baruch Fischhoff^{1,2}, Alexander Davis¹ and Stephen B Broomell²¹ Department of Engineering & Public Policy, Carnegie Mellon University, USA² Department of Social & Decision Sciences, Carnegie Mellon University, USA³ Author to whom any correspondence should be addressed.E-mail: barrydewitt@cmu.edu**Keywords:** decision-making, risk perception, risk communication, signal detection theory, multidimensional scalingSupplementary material for this article is available [online](#)**Abstract**

Lay judgments of environmental risks are central to both immediate decisions (e.g., taking shelter from a storm) and long-term ones (e.g., building in locations subject to storm surges). Using methods from quantitative psychology, we provide a general approach to studying lay perceptions of environmental risks. As a first application of these methods, we investigate a setting where lay decisions have not taken full advantage of advances in natural science understanding: tornado forecasts in the US and Canada. Because official forecasts are imperfect, members of the public must often evaluate the risks on their own, by checking environmental cues (such as cloud formations) before deciding whether to take protective action. We study lay perceptions of cloud formations, demonstrating an approach that could be applied to other environmental judgments. We use signal detection theory to analyse how well people can distinguish tornadic from non-tornadic clouds, and multidimensional scaling to determine how people make these judgments. We find that participants ($N = 400$ recruited from Amazon Mechanical Turk) have heuristics that generally serve them well, helping participants to separate tornadic from non-tornadic clouds, but which also lead them to misjudge the tornado risk of certain cloud types. The signal detection task revealed confusion regarding shelf clouds, mammatus clouds, and clouds with upper- and mid-level tornadic features, which the multidimensional scaling task suggested was the result of participants focusing on the darkness of the weather scene and the ease of discerning its features. We recommend procedures for training (e.g., for storm spotters) and communications (e.g., tornado warnings) that will reduce systematic misclassifications of tornadicity arising from observers' reliance on otherwise useful heuristics.

'If you hear a roaring sound, or see a funnel cloud, you should seek shelter immediately.'—Canadian Broadcasting Corporation (CBC) Radio, 1 July 2014, 2:00pm EST

More than 1000 tornadoes hit the United States and Canada each year, resulting in deaths and property damage, sometimes on a catastrophic scale (National Oceanic and Atmospheric Association (NOAA) 2014a).

For example, the May 2011 tornado in Joplin, Missouri, resulted in 159 deaths and caused nearly USD 3 billion in damage (Masters 2012, NOAA 2015). In Europe, all but a handful of countries have experienced tornadoes (Groenemeijer and Kuhne 2014), with many European nations currently improving their severe storm forecasting systems (Rauhala and Schultz 2009).

Field reports of severe storms are essential to tornado warnings and forecasting because remote sensing data cannot definitively identify some critical storm features (Doswell *et al* 1998, League *et al* 2010, NOAA 2011, Brotzge and Donner 2013). These forecasts are limited

by the computational challenges of data processing, insufficient observations of high spatiotemporal resolution, and the lack of valid radar- or satellite-visible predictors for some tornadic weather patterns (Brotzge and Donner 2013). Moreover, there are too few trained storm spotters to reliably provide warnings with the lead time needed for effective action (Rauhala and Schultz 2009). As a result, members of the public must often evaluate the risks on their own (Doswell *et al* 1998, Durage *et al* 2012, Brotzge and Donner 2013). Their success depends on how well they can gauge the risks from observable cues (Brotzge and Erickson 2010, Lindell and Perry 2012, Brotzge and Donner 2013). As in the CBC broadcast, valid cues can be both visual (e.g., clouds, wind, hail) and auditory (e.g., the 'roar of a 'freight train' ') (NOAA 2011, Lindell and Perry 2012, p 168).

Weather forecasters' communications attempt to help members of the public take appropriate safety measures. Research has found that people often seek visual and auditory cues before deciding whether to heed them (Liverman and Wilson 1981, Sorensen 2000, Lindell *et al* 2005, Dash and Gladwin 2007, League *et al* 2010, Lindell and Clayton 2012). For example, the sight of rising water levels has been found to prompt residents to evacuate before an approaching hurricane (Morss and Hayden 2010). However, relatively little is known about how and how well laypeople extract information from most environmental cues. Here, we offer a general approach to addressing this question, within the normative–descriptive–prescriptive framework of behavioural decision research (Edwards 1954, Fischhoff and Kadvanly 2011), illustrated with lay evaluation of visual cues for tornado risk. We begin by briefly summarizing meteorological research into the diagnostic value of those cues (normative analysis). We then assess lay observers' ability to extract diagnostic information from pictures of clouds, based on methods from decision science (descriptive analysis). These results then frame recommendations for policies designed to accommodate and reduce the limits to lay abilities (prescriptive analysis). In addition to advancing the understanding of tornado risks, we hope to illustrate a method applicable to other environmental hazards, both immediate (e.g., floods) and long-term (e.g., living in an area prone to flooding).

The training of storm spotters is based on meteorologists' analysis of the cues most useful for predicting severe thunderstorms and tornadoes. US National Weather Service (NWS) training emphasizes visual cues (NOAA 2011), such as *wall clouds*, a characteristic lowering of the storm from which tornadoes descend, and having a 'solid or hard-looking storm tower with a cauliflower appearance,' indicating strong storm activity (NOAA 2011, p 22). Figure 1 shows examples, all taken from official NWS materials or from professional weather photographers, sources that included verification of the weather scene in the photograph (see methods). Figure A1 in supplementary information (A)

(SI(A)) provides further examples. Unfortunately, no cue is perfectly predictive and the most predictive ones, such as the formation of a funnel cloud (figure 1(F)), occur so late that they provide little response time. Moreover, some important cues can be difficult to distinguish from less tornadic phenomena (NOAA 2011, NWS 2014b). For example, *shelf clouds* (figure 1(C)) may look very similar to wall clouds. Although storm spotters are trained to make these distinctions (NOAA 2011, NWS 2014b), even they sometimes have difficulty (NOAA 2011, Brotzge and Donner 2013, NWS 2014b). SI(A) describes more fully the tornado detection and warning procedures and the performance metrics that provide the normative analysis underlying the present research. In this study, we focus on North America, where tornadoes pose the greatest threat and where tornado forecasting is most developed; see Rauhala and Schultz (2009) for a comparison of the American and European systems.

Here, we present descriptive research into how and how well laypeople detect tornado danger from clouds, applying two approaches from *psychophysics*, the study of psychological responses to physical stimuli: signal detection theory (SDT) (Green and Swets 1988, Wickens 2001, Macmillan and Creelman 2004) and multidimensional scaling (MDS) (Baird and Noma 1978, Borg and Groenen 2005, Borg *et al* 2012). SDT separates decision-maker ability (called *sensitivity*) to detect signals (here, tornado potential) from the decision rule for acting as though a signal exists. Thus, SDT recognizes that false alarms (FAs) for tornadoes depend on both how well forecasters can detect them and how cautious forecasters want their forecasts to be (Harvey *et al* 1992, Brooks 2004). Similarly, SDT can distinguish between situations where people fail to seek shelter (a) because they do not see the risk, and (b) because their belief about the odds of a tornado occurring is incorrect or because the inconvenience of a FA feels too high. MDS extracts the dimensions that decision-makers use in determining the extent to which stimuli are similar (e.g., how much cloud formations look like an archetype of a tornado), and can reveal perceptual processes that individuals may not realize or be able to articulate (e.g., 'It just looks like a tornado.'). These methods formalize the analysis of uncertain situations.

We use separate, but interrelated, tasks for SDT and MDS. The SDT task asks participants to judge whether each of 50 pictures of clouds drawn from public sources⁴ was taken when a tornado watch was in effect. We elicited a categorical judgment, rather than a continuous one, given the meteorological difficulty of assigning degrees of tornadicity to cloud formations. We chose 'tornado watch' as a relatively

⁴ A library of all the stimuli collected during the creation of this experiment is available at <http://goo.gl/WXQPuW>. A subset of the library was used in the experiment (as indicated in the library).

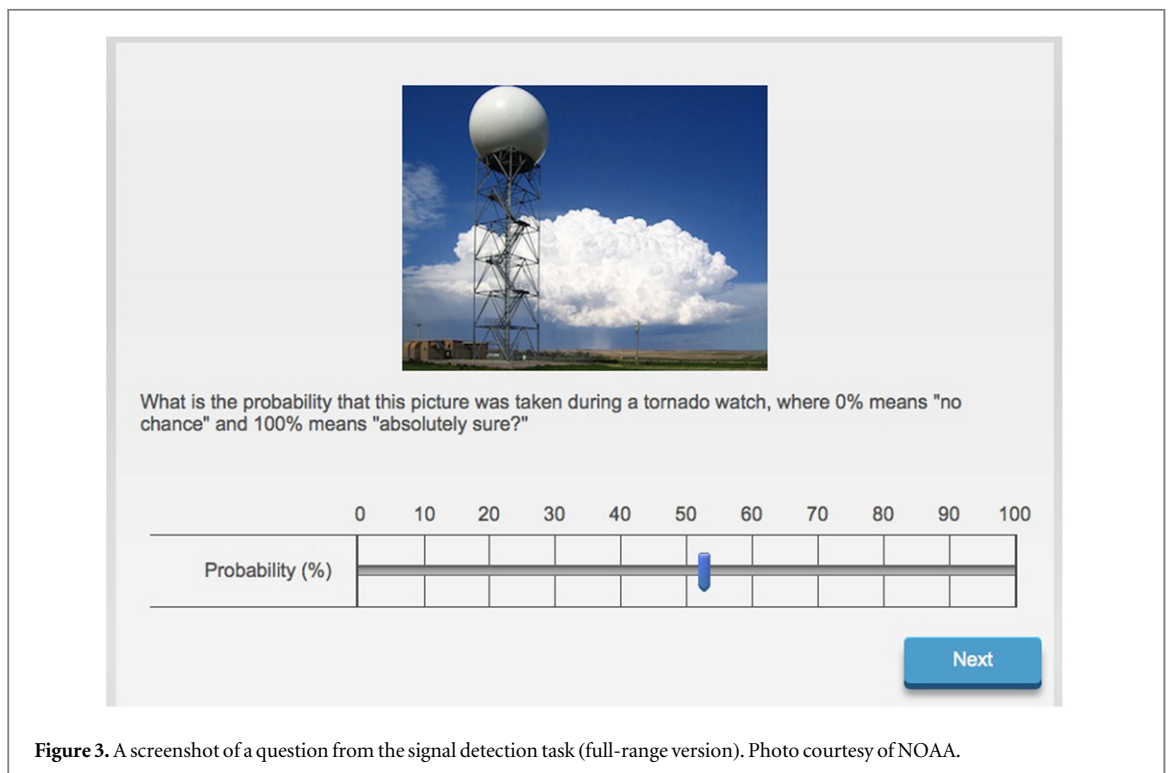
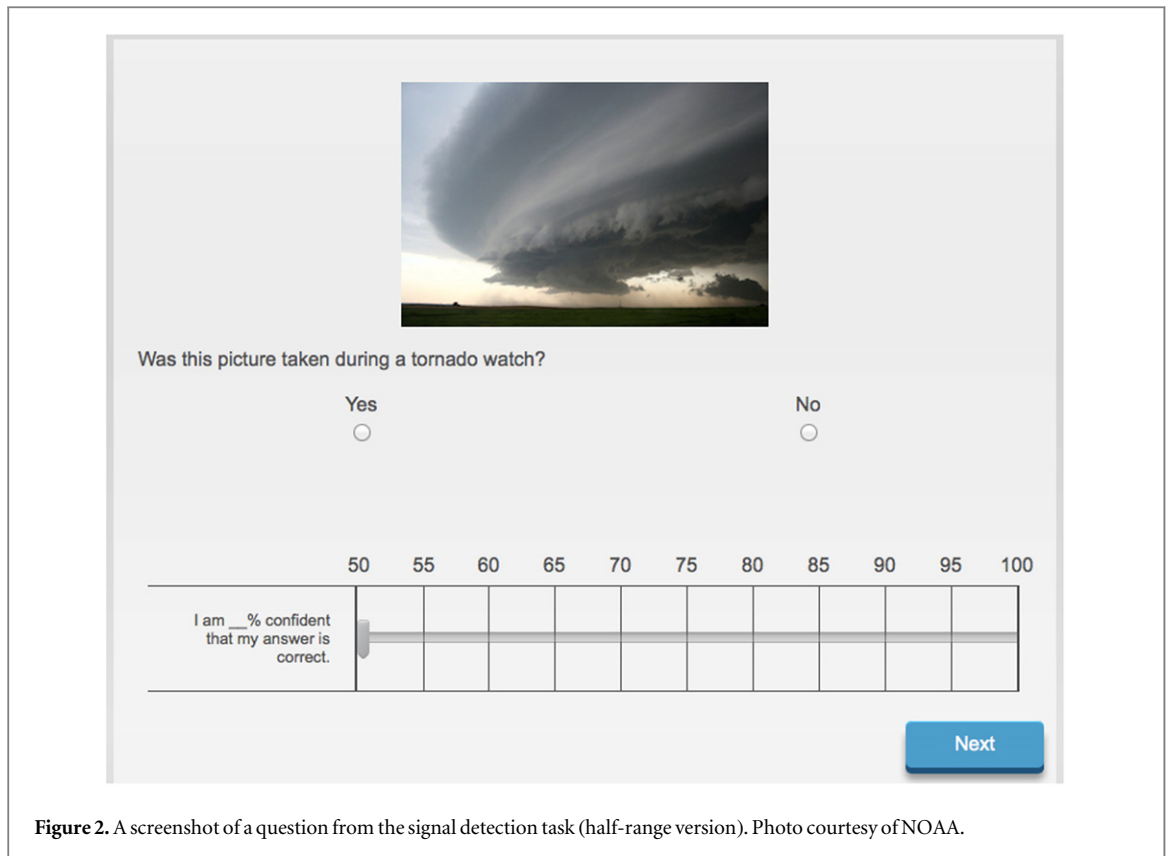


familiar, and important, category; see discussion for more details. The MDS task asks participants which picture seems more tornadic, in pairs involving 11 of the 50 pictures. All participants performed both tasks, with half starting with SDT and half with MDS. We expected that performing the MDS task first would improve SDT performance as a result of providing an opportunity to study some of the stimuli (Mundy *et al* 2007, de Zilva and Mitchell 2012). On the other hand, we did not expect the SDT task to affect MDS judgments, thereby producing an asymmetrical transfer effect, whereby one task affects the other but not vice versa (Poulton and Freeman 1966).

Our experimental design also compares two probability response modes for the SDT task (Lichtenstein *et al* 1982). The *half-range* task has participants decide whether a picture was taken during a tornado watch,

and then give their confidence in that judgment with a probability in the (50%–100%) range. The *full-range* task has participants give a probability judgment for the picture having been taken during a tornado watch, using the (0%–100%) range. We expected the half-range response mode to elicit slightly better sensitivity, by encouraging greater focus on discrimination in its first stage.

Understanding the effects and nuances of tornado experience is a topic of interest in natural hazards research (Blanchard-Boehm and Cook 2003, Nagele and Trainor 2012, Drost 2013, Demuth 2014). As a result, we also collected information about participants' tornado experience, weather knowledge, emergency preparedness, numeracy, and demographic characteristics, as covariates for modelling performance in the SDT task.



Methods

Participants ($N = 400$) recruited via Amazon Mechanical Turk completed three tasks: SDT evaluation of cloud tornadicity, MDS comparison of cloud similarity, and personal information questions. Participants

were randomly assigned to one of four cells in a 2×2 design, with the factors of (a) SDT response range (half-range or full-range) and (b) task order (SDT or MDS first). Details appear in SI(B).

The SDT and MDS tasks use photographs of clouds taken from public sources. Each picture was

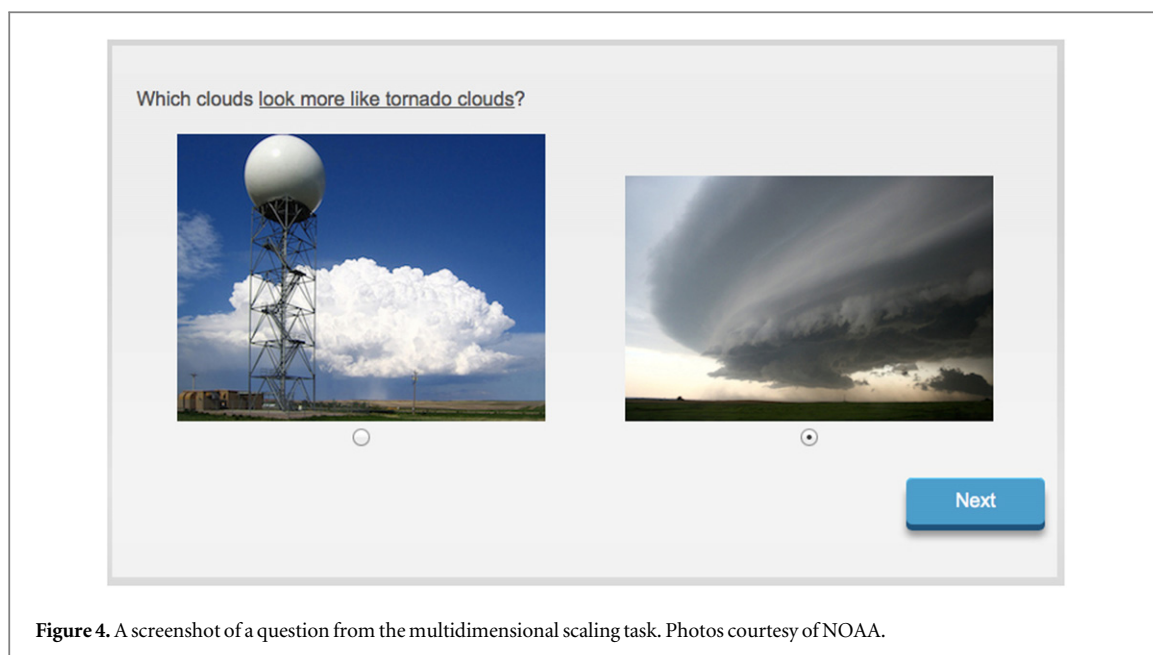


Figure 4. A screenshot of a question from the multidimensional scaling task. Photos courtesy of NOAA.

classified as tornadic or non-tornadic, based on its description by its source and NWS criteria (NOAA 2014b, 2014c). There were 25 pictures of each type.

Figures 2 and 3 show examples of the two SDT tasks, half- and full-range, differing solely in how responses were elicited. The ‘tornado watch’ formulation of both prompts was designed to elicit a judgment of whether a cloud formation was tornadic or non-tornadic, which we then compared with its actual classification based on the NWS criteria. Participants saw 50 pictures, without being told the 50% base rate of tornadic stimuli. Three of the 50 pictures were exceptionally easy (two *Wizard of Oz*-type tornadoes, one clear blue sky), in order to assess whether participants were paying attention (see figure 1). The two tornadoes opened and closed the task. The other 48 stimuli (including the clear blue sky) were in a random order for each participant. Response time data were collected in order to see if people spent more time in one condition than another.

After completing these judgments, participants answered three open-ended questions:

- In 25 words or less, what does a ‘tornado watch’ mean to you?
- What made cloud pictures look like they were taken during a tornado watch?
- What made cloud pictures look like they were not taken during a tornado watch?

In the MDS task, participants saw all pairwise combinations of 11 pictures drawn from the full set of 50, yielding 55 pairs. These 11 pictures were chosen to create pairs varying widely in similarity, so that the

MDS algorithms could more easily uncover the psychological dimensions underlying participant judgments (Borg *et al* 2012); 6 were tornadic and 5 non-tornadic. See figure 4.

After completing these tasks, participants were asked for their age, gender, education, and state/province of residence. They also answered questions about their tornado experience, interest and knowledge regarding weather, and level of emergency preparedness. Finally, they completed two standard instruments: (a) the seven-item weather salience questionnaire (short form), designed to gauge the psychological significance of weather (Stewart 2009, Stewart *et al* 2012); and (b) the eight-item Subjective Numeracy Scale (Fagerlin *et al* 2007), which has been found to predict individuals’ ability to understand risk communications (Zikmund-Fisher *et al* 2007). Table 1 provides some example questions.

Results

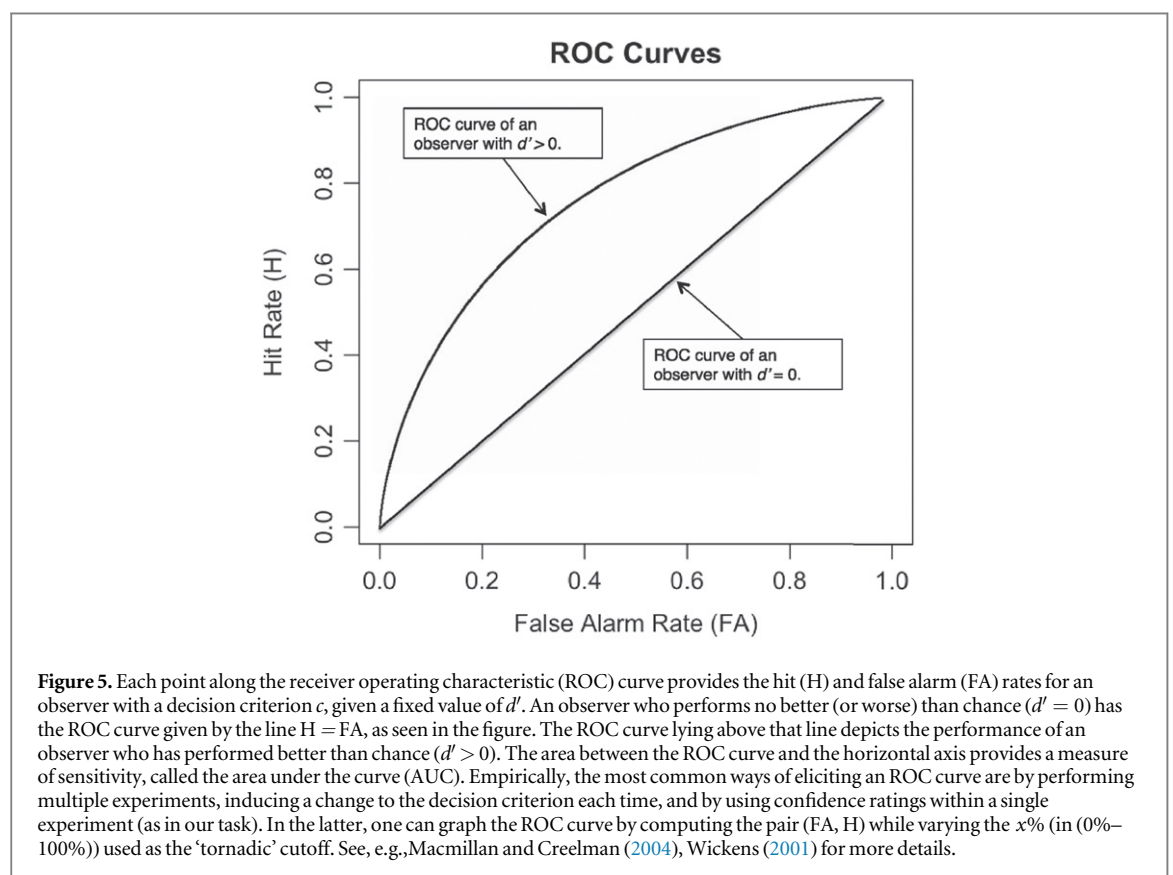
After describing the sample’s demographic characteristics, we present results for the SDT and MDS tasks. Finally, we analyse individual-level predictors of SDT performance.

Sample demographics

The sample had 156 men and 244 women, whose age ranged from 18–75 years old (mean = 37, median = 33). Three reported ‘some high school’ as their highest level of education, 48 ‘high school,’ 20 ‘community college,’ 130 ‘some college,’ 143 ‘college,’ 47 ‘graduate school,’ and 9 ‘professional school.’ Participants were widely distributed geographically, allowing for varied tornado experiences (as reported below).

Table 1. Example personal information questions. See SI(E) for the full list of questions.

Topic	Example items	Response options
Weather salience	'I take notice of changes that occur in the weather.'	Never (1), seldom (2), sometimes (3), usually (4), all of the time (5).
	'I notice how the clouds look during various kinds of weather.'	Never (1), seldom (2), sometimes (3), usually (4), all of the time (5).
Subjective numeracy	How good are you at working with fractions?	Six-point scale (1–6), from 'not at all' to 'extremely.'
	How good are you at working with percentages?	Six-point scale (1–6), from 'not at all' to 'extremely.'
Experience	Have you ever experienced a tornado directly?	Yes (1), no (0).
	Do you have friends or family who have experienced a tornado directly?	Yes (1), no (0).
Weather interest	How knowledgeable are you about the weather?	Four-point scale (1–4), from 'not at all' to 'extremely.'
	How interested are you in the weather?	Four-point scale (1–4), from 'not at all' to 'extremely.'



Twenty-two participants failed at least one of the attention-check questions. All analyses were run with those individuals, and we make note of any substantive results that would change if those participants were excluded.

Signal detection results

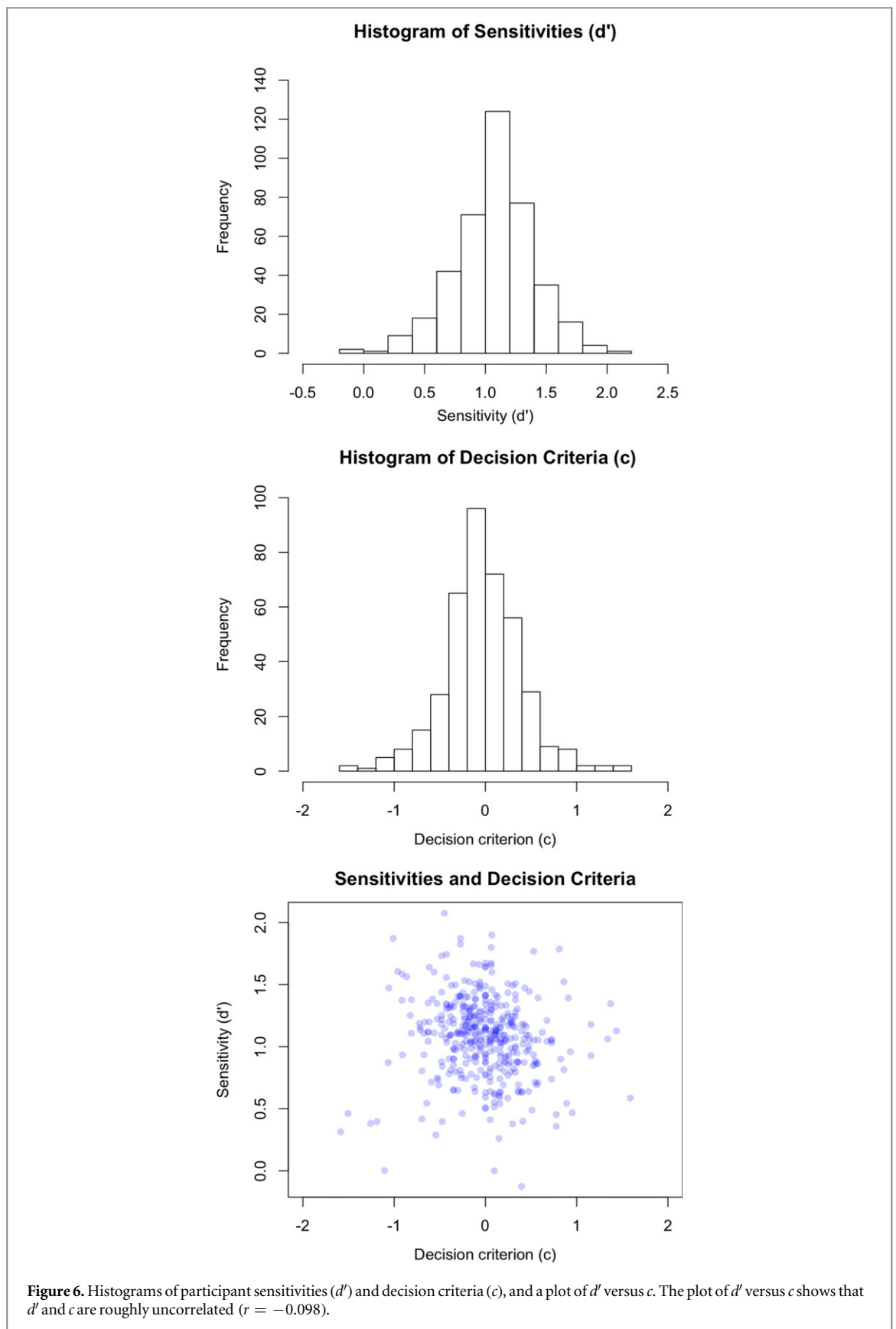
Sensitivity (d')

Signal detection parameters were estimated using participants' choices (i.e., tornadic or non-tornadic) and confidence judgments. In the half-range mode, those responses were used directly. In the full-range mode, 50% was taken as the cutoff (randomly

treating 50% responses as tornadic or non-tornadic) and the absolute deviation from 50% was used as the measure of confidence (following Juslin *et al* 1997, who provide a normative analysis of this procedure)⁵.

In order to estimate sensitivity (d'), we first computed the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for each participant; see figure 5. AUC is bounded between 0 and 1. To create an unbounded measure, which can be

⁵ This convention allows converting half-range responses to full-range, and vice versa.



compared to other methods for estimating d' (see SI (C)), we also computed $d' = z(\text{AUC}) \cdot \sqrt{2}$, where z is the quantile function of the normal distribution with mean 0 and variance 1. See SI(C) and Wickens (2001) for details.

Figure 6 shows participant d' s. The sample mean d' is 1.08, with a 95% confidence interval for the mean of (1.04, 1.11). Thus, most participants performed moderately well, with that mean indicating a classification accuracy of about 70% (for individuals trying to

maximize the number of correct classifications) (Lynn and Barrett 2014).⁶

Bias (c)

The *decision criterion* (also called the *decision bias*, *decision rule*, or *response bias*) for each participant was estimated from that individual's hit (H) and FA rates⁷. There are many measures of decision bias. We use⁸ $c = -0.5(z(H) + z(FA))$.

Normatively, c is a function of both the observer's belief about the base rate of the signal, and their payoff matrix (i.e., their valuation of Hs, misses, correct rejections, and FAs) (Coombs *et al* 1970, Baird and Noma 1978). If participants value the outcomes such that they care only about their classification accuracy, c should reflect their belief about the base rate of the signal, as they should cite the more common category whenever they cannot make a discrimination. Alternatively, participants might differentially value Hs, misses, correct rejections, and FAs, in which case c will reflect that bias.

We chose c as a measure of bias because it has several attractive properties (Macmillan and Creelman 2004, ch 2), including its easy interpretation. If c reflects the observer's valuation of Hs, misses, correct rejections, and FAs, then $c = 0$ (called a 'neutral' decision rule) indicates that signal (tornadic) and noise (non-tornadic) stimuli are of equal importance to the observer, positive values (called 'conservative') indicate less tolerance for FAs relative to misses, and negative values (called 'lax') indicate greater tolerance for FAs relative to misses (Baird and Noma 1978, Wickens 2001, Macmillan and Creelman 2004). If c solely reflects beliefs about the base rate of the signal, analogous interpretations exist for neutral, conservative, and lax criteria: a neutral criterion implies a perceived base rate of 50%, a conservative criterion implies the perception that the signal rarely occurs, and a lax criterion implies the perception that the signal is common.

The mean c is -0.02 , with a 95% interval of $(-0.06, 0.02)$. Figure 6 shows a histogram of participant c s. On average, participants had an approximately neutral decision criterion, but with a range of individual policies. Figure 6 also contains a plot of d'

versus c , showing that d' and c are roughly uncorrelated ($r = -0.098$).

According to a strict utility interpretation, the positive c values among the distribution of c s in figure 6 would indicate that many participants wanted to avoid FAs at the cost of increasing the number of misses, which seems unlikely given the dangers posed by tornadoes. Instead, we think the diversity of c values observed could be better explained by participants attempting to maximize their accuracy, setting their c s to reflect their beliefs in the odds of a tornadic stimulus occurring in the task. Similarly, it is likely that the nature of the task—asking for a tornadicity classification, rather than, say, about what the participant would do given the weather scene—supports a probability interpretation over a utility interpretation. Thus, we shall interpret c as reflecting participants' priors, recognizing that some participants may also have incorporated utilities into their decision criteria.

Stimulus-level results

Figure 7 shows how accurately the individual stimuli were classified as tornadic (signal) or non-tornadic (noise), based on the mean full-range probability assigned (using all responses, converting those given for the half-range task to probabilities on the 0%–100% scale; see footnote 5). As can be seen, although the mean probability judgment was higher than 50% for most tornadic clouds (labelled with an 's') and lower for most non-tornadic ones (labelled with an 'n'), there were some exceptions⁹. The MDS scaling results guided our interpretation of the features affecting the accuracy of these judgments.

Multidimensional scaling

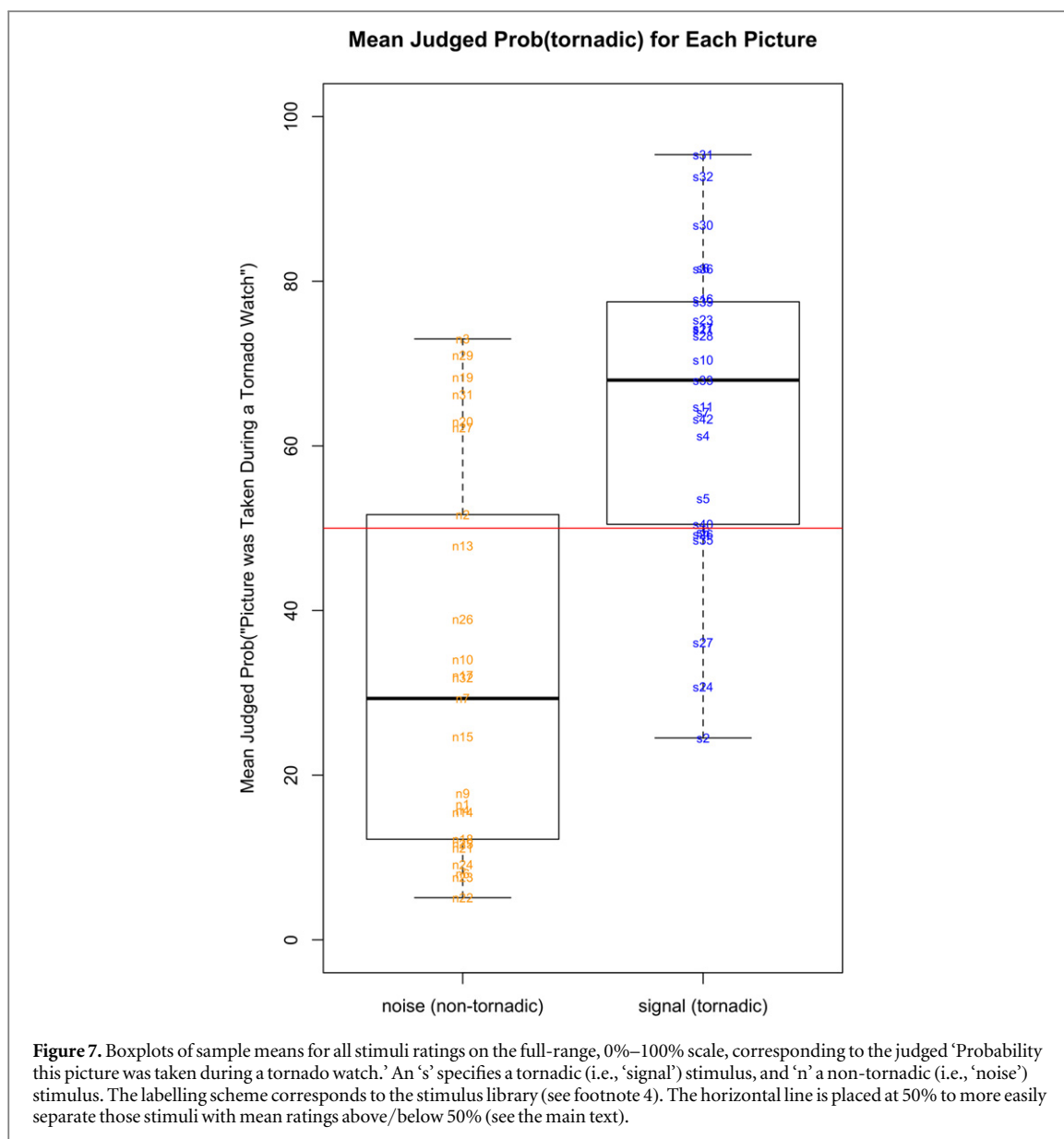
MDS creates a geometric representation of the stimuli using participants' judgments of the more tornadic picture among each of the 55 paired comparisons created from the 11 selected pictures. Using standard criteria (see SI(D)), we found the best fit with the two-dimensional representation in figure 8, with dimensions that we interpret as reflecting (a) the darkness or 'ominousness' of the picture and (b) its clarity, referring to how easy it is to discern and interpret the features of the cloud formation that mark it as tornadic or non-tornadic. Looking at the specific pictures, we find potential sources for misclassifications: pictures of (non-tornadic) shelf clouds (1, 3) and (tornadic) supercells (6, 7, 8) were seen as similar. Pictures with upper- and mid-level tornadic cloud features (9, 10) were seen as similar to non-tornadic fair weather cumulus clouds (2). Among the tornadic photographs,

⁶ Changes in the NWS criteria used to classify the stimuli as tornadic or non-tornadic could change based on advances in the meteorological understanding of tornadoes. While such a change would necessarily affect our estimates of the signal detection parameters, it would not impact our conclusions. Any reader planning to use our values as estimates of the population's psychophysical parameters should be aware of this uncertainty. We thank an anonymous reviewer for highlighting this point.

⁷ The AUC calculation, used to estimate d' , is based on an ROC curve, which shows H and FA rates associated with alternative decision rules. As a result, it provides no measure of bias.

⁸ In order to account for H or FA equalling 1 or 0, we added 0.5 to each cell of the contingency table. This converts, for example, an H of 1 (i.e., all 25 tornadic photographs correctly classified as tornadic) into the fraction $\frac{25.5}{26}$, and similarly, a FA of 0 into $\frac{0.5}{26}$ (Hautus 1995, Macmillan and Creelman 2004).

⁹ Note that, in figure 7, some of the tornadic stimuli below 50% (s6, s26, s35) are very close to the 50% mark, as is the dust storm (n2) that is above 50%. Similarly, a supercell (s40) and a wave cloud (n13) are very close to the 50% mark, though these are above and below 50%, respectively. In fact, if we consider a stimulus misclassified if the 95% confidence interval for its mean rating contains 50%, then s40 and n13 are misclassified as well.



9 and 10 were seen as the most dissimilar to the tornadic extreme represented by 11. A unidimensional scaling (provided in SI(D)) confirmed the upper- and mid-level tornadic clouds are likely to be confused with non-tornadic clouds. Put simply, shelf clouds look worse than they are because they are both ominous and dark, with structural features that might seem unfamiliar to the untrained eye. Clouds with upper- and mid-level tornadic features are slightly unusual, but the blue skies that surround them mean they are fairly bright, and (we speculate) less ominous to the layperson. These results were not affected by whether participants performed the MDS task first or the SDT task first.

The MDS results shed light on the systematic misclassifications that occurred in the SDT task (figure 7). All of the shelf clouds (n3, n19, n20, n27, n29) were classified inaccurately, as were four of the five with upper- and mid-level cloud features (s2, s6, s26, s27). The other tornadic pictures that were misclassified

contained *mammatus* clouds (s24, s35): the MDS would suggest this was because of their brightness. A picture of lightning (n31) and a dust storm (n2) were misclassified as tornadic: the former is certainly ominous, and the latter is dark, with features that are (almost by definition) difficult to discern.

We directly tested the link between the MDS results and SDT results by regressing the mean SDT full-range scores of the 11 common stimuli onto their coordinates in the two-dimensional MDS space. Table 2 displays the estimated coefficients for the darkness and clarity dimensions, which account for a large proportion of the variance ($R^2 = 0.97$). We see that the darkness/ominousness dimension is clearly the more important dimension, and that the linear regression does a good job of modelling the SDT scores, as would be expected if MDS captures the attributes underlying judgments of tornadicity. We also regressed each individual's SDT scores on the MDS coordinates, to see how well individual SDT scores

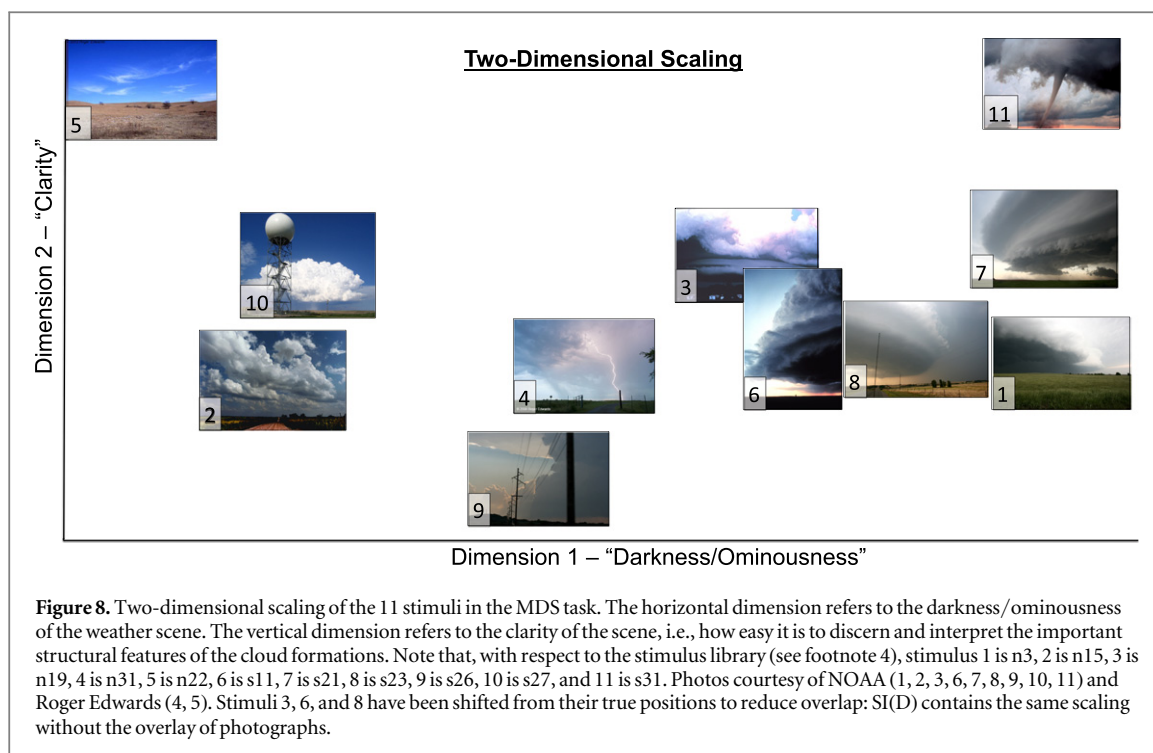


Table 2. Coefficient estimates and standard errors for a regression modelling the mean SDT full-range scores (i.e., mean judged *prob(tornadic)*) as a function of the MDS coordinates.

	Dependent variable: prob(tornadic)
Darkness/ominousness	25.89 ^a (1.57)
Clarity	1.05 (1.57)
Constant	4.61 (4.01)
Observations	11
R ²	0.97
Adjusted R ²	0.96
Residual std. error	4.94 (df=8)
F Statistic	136.51 ^a (df=2; 8)

Note.

^a $p < 0.01$.

were modelled by the aggregate MDS solution. The average R^2 was 0.7. Note that to make interpretation easier, the coordinates of the MDS representation were standardized, and the space transformed so that its origin represents the location of a hypothetical stimulus having essentially no ‘darkness/ominousness’ and being the least ‘clear’ possible.

Individual difference measures

Participants’ responses to the seven questions of the weather salience questionnaire (short form) were summed (with question 5 reverse scored) to form a

weather salience score (WxSQ), which had a Cronbach’s α of¹⁰ 0.61. Participants’ responses on the subjective numeracy questions were averaged to produce a subjective numeracy score (SNS), which had a Cronbach’s α of 0.86. We created a Tornado or Dixie Alley score (0, 1), reflecting whether participants were in one of two tornado-prone regions. The former includes Texas, Oklahoma, Kansas, Missouri, Iowa, Nebraska and South Dakota ($n = 54$) and the latter Louisiana, Arkansas, Mississippi, Alabama, Tennessee, and Georgia¹¹ ($n = 35$). The three questions regarding knowledge, interest, and forecast-following behaviour were transformed to a 1–4 scale, and then averaged (and standardized) to create a (self-reported) *meteorophily* score, which had a Cronbach’s α of 0.68. The nine items related to participants’ severe-storm experience were summed (and then standardized) to create an *experience* score, which had a Cronbach’s α of 0.76. The item measuring the *impact* of participants’ tornado experience was kept separate and standardized. Having emergency supplies, a planned place to take shelter, and knowing that a tornado warning is more serious than a watch—all binary variables—were kept separate and untransformed. Finally, we created a binary variable *fail*, equal to 1 if the participant failed any of our attention checks, and 0

¹⁰ Cronbach’s α is a measure of internal consistency (Cronbach 1951). Bland and Altman (1997) provide a short and accessible introduction to Cronbach’s α .

¹¹ Though there are no formal definitions for these areas, we used those from NOAA: see www.ncdc.noaa.gov/climate-information/extreme-events/us-tornado-climatology/trends.

Table 3. Coefficient estimates and standard errors for regressions modelling d' and c , based on the individual difference measures and the experimental conditions.

	Dependent variable:	
	d'	c
Order (0 = MDS-SDT)	−0.02 (0.04)	0.17 ^b (0.06)
Mode (0 = full-range)	−0.05 (0.04)	−0.05 (0.06)
Experience	0.07 ^b (0.02)	0.01 (0.03)
Impacts	−0.05 ^b (0.02)	0.01 (0.03)
Alley (1 = lives in Tornado/Dixie Alley)	0.02 (0.04)	−0.0002 (0.05)
Meteorophily	−0.003 (0.02)	−0.003 (0.03)
Shelter (1 = has planned shelter)	0.08 ^a (0.04)	−0.10 ^a (0.05)
Supplies (1 = has supplies)	−0.05 (0.03)	−0.008 (0.05)
Watch-warning (1 = knows warning is more serious)	0.04 (0.04)	0.03 (0.06)
WxSQ (weather salience)	0.01 (0.005)	−0.02 ^b (0.007)
SNS (subjective numeracy score)	0.03 ^a (0.01)	−0.01 (0.02)
Fail (1 = failed attention checks)	−0.45 ^b (0.07)	−0.07 (0.09)
Order × mode	0.03 (0.06)	−0.11 (0.08)
Constant	0.75 ^b (0.14)	0.63 ^b (0.20)
Observations	400	400
R^2	0.20	0.10
Adjusted R^2	0.17	0.07
Residual Std. Error (df = 386)	0.30	0.41
F Statistic (df = 13; 386)	7.40 ^c	3.38 ^c

Note.

^a $p < 0.05$.

^b $p < 0.01$.

otherwise. SI(E) further describes the creation of these measures.

Modelling SDT parameters

Table 3 predicts d' and c as the dependent variables, using the individual difference measures and the experimental condition factor variables (SDT response mode and task order) as covariates. See SI(F) for regression diagnostics.

The most striking results are the higher values of d' for participants who report having had more personal experience with tornadoes and having done more to prepare for them, and the lower values of d' among those who report greater impacts from tornadoes and those who failed the attention checks. None of the

following variables predicted greater sensitivity (d'): living in Tornado or Dixie Alley¹², finding weather more salient (WxSQ), reporting greater meteorophily, or knowing that a tornado warning is more serious than a tornado watch. Decision bias (c) was related to two individual difference measures: participants with higher WxSQ scores (indicating that weather was more salient for them) were more likely to see danger in the cloud formations¹³, as were participants who reported having planned a place to shelter during severe weather. However, the finding about sheltering was no longer significant when participants who failed any of the attention-checks were excluded.

Because of the interaction in the model, the coefficients on *order* and *mode* in table 3 show simple effects (i.e., the effect of task order and of response mode when the other variable is set to 0). We used the results in table 3 and the variance–covariance matrix of the coefficients to investigate main effects as well. We found that d' was unrelated to whether the SDT task came first or second, and whether the SDT task used the half- or full-range mode. However, participants were significantly less likely to categorize stimuli as tornadic (i.e., had higher c values) when they completed the SDT task first, and significantly more likely to categorize stimuli as tornadic (i.e., had lower c values) when responding using the half-range mode. The estimates for the main effect of task order and response mode on c were 0.12 and -0.11 , respectively, with 95% confidence intervals of (0.04, 0.20) and (-0.19 , -0.03).

Discussion

Psychophysical performance

The present study provides the first psychophysical investigation of tornado risk perception. Our signal detection analysis found that participants had some ability to distinguish between tornadic and non-tornadic stimuli, with the mean d' of 1.08 equating to about 70% accuracy for individuals with the roughly neutral decision rule found here ($c = -0.02$). The low correlation between d' and c (figure 6) shows that participants' decision criteria are independent of their perceptual ability to detect tornadic weather.

With respect to individual differences in psychophysical performance (table 3), participants who reported more experience with tornadoes demonstrated greater discrimination ability, whereas those who reported greater impacts from tornadoes demonstrated less and living in a tornado-prone area had no effect. However, d' was higher for individuals who

¹² A secondary analysis found that living in Tornado Alley predicts a 0.1 increase in d' , while living in Dixie Alley predicts a 0.1 decrease in d' . The two groups combined, however, had only 89 people, from the sample of 400.

¹³ We thank an anonymous reviewer for suggesting this probability-based interpretation.

reported having planned a place to shelter, who rated themselves higher on the SNS, and who passed our attention checks. Decision bias (c) was associated with only two individual difference measures. Participants with a higher weather salience score (WxSQ) were more likely to categorize cloud formations as tornadic, as were those who reported having planned a place to shelter.

Discrimination was the same with the half-range and full-range SDT tasks, contrary to our prediction that the comparison required by the half-range task would lead people to look more closely. Participants did, indeed, spend an average 1.25 s longer on it per judgment; however, that time apparently went into completing the task, rather than thinking about its content. For c , participants using the half-range mode had a more lax criterion (i.e., were more likely to categorize a stimulus as tornadic). We had no prediction regarding the effect of response mode on c , but speculate that explicitly choosing the category in the half-range mode made utilities more salient than in the full-range task, where the category was inferred from the confidence rating. As a separate methodological point, the best single predictor of d' was the attention-check indicator variable. Participants who improperly classified any one of the two funnel clouds or the clear blue sky showed poorer discrimination ability on the other 47 stimuli, supporting the utility of such checks (Oppenheimer *et al* 2009, Downs *et al* 2010).

We also found no better discrimination among participants who did the MDS task first, contrary to our prediction that examining a subset of the stimuli would improve performance. However, d' calculated for just the 11 stimuli used in the MDS task was higher for those who did the MDS task first (mean of 0.63 versus 0.56, although the 95% confidence interval for the difference in means, $(-0.02, 0.17)$, includes no difference). For related results, see de Zilva and Mitchell (2012), Lavis and Mitchell (2006), Mundy *et al* (2007), who investigate the features of prior exposure that improve discrimination of stimuli and memory for their attributes. Participants who performed the MDS task first were more likely to classify stimuli as tornadic in the SDT task (had lower c values), suggesting that prior exposure to stimuli led to an increased sense of tornadic risk.

The MDS results suggest heuristics that might affect performance—in the SDT task and perhaps in life. Specifically, associating darkness with tornado risk may lead to viewing upper- and mid-level tornadic clouds (the visible blue sky) as less tornadic than they are; such clouds require some sophistication to notice and understand the features that mark them as dangerous, the lack of which could also lead to misclassification. Regression showed that the darkness/ominousness and clarity dimensions effectively modelled the mean response in the SDT task for the 11 stimuli included in both the MDS and SDT tasks. It also showed that darkness/ominousness was much more

important than clarity; similarly, additional MDS analyses (see SI(D)) found that the darkness/ominousness dimension was more important than clarity in explaining judgments of cloud similarity. As cues of actual (rather than perceived) tornadicity, darkness/ominousness has validity, given that severe storms are dark when they are overhead or very close by. On the other hand, cloud formations with more easily discernible features (i.e., higher on the clarity dimension) are, to the best of our knowledge, no more or less likely to be tornadic¹⁴. Future work could, by scaling a large set of stimuli, clarify both the link between the dimensions laypeople are using and their validity as markers of tornadic threat, and the difficult-to-define 'clarity' dimension.

Our approach classified stimuli dichotomously, as tornadic or non-tornadic. An alternative approach is to characterize the stimuli by their probability of resulting in a tornado. However, we could not find those probabilities in the literature, nor the observational data with which to construct them ourselves. By using the NWS storm-spotter training curriculum as a classification guide, we have relied on expert opinion to decide what is tornadic.

Participants were prompted to classify the stimuli with a question that asked about the occurrence of a tornado watch (figures 2 and 3). We considered many alternatives to that chosen formulation, including: 'Are these clouds associated with tornadoes?', 'Are these clouds often associated with tornadoes?', 'Are these clouds tornado clouds?', 'Will these clouds form a tornado?', with appropriate changes to other instructions. While recognizing that some (more sophisticated) participants might have preferred a less deterministic formulation, we saw no tractable way of expressing 'associated' so that it would be interpreted similarly by all respondents (Wallsten *et al* 1986). One concern with our chosen 'watch' formulation was whether participants would correctly interpret the prompt. Notwithstanding possible concerns about self-reports (Ericsson and Simon 1980), pre-testing found that participants generally interpreted it as desired, and the open-ended questions asked after the SDT task found that participants in the actual experiment generally did so as well.

Policy implications

Our results suggest that lay observers have generally effective heuristics for deciding whether clouds are tornadic. However, those heuristics lead to biases,

¹⁴ The validity of the dimensions can also be investigated by regressing the actual classifications (tornadic or non-tornadic) of the 11 stimuli in the MDS task on their coordinates in the MDS space. This logistic regression finds that, for a 1-unit increase on the darkness dimension, the odds of a stimulus being tornadic are 2.2 times greater. (The coordinates have been standardized, thus 1-unit corresponds to one standard deviation.) For a 1-unit increase on the clarity dimension, the odds of a stimulus being tornadic are 1.1 times greater. However, these parameters cannot be precisely estimated due to the small sample size.

including systematically misclassifying shelf clouds as tornadic, clouds with upper- and mid-level tornadic features (such as well-defined storm towers) as non-tornadic, and mammatus clouds as non-tornadic. Looking at the MDS representation (figure 8), projecting onto the horizontal dimension produces a mostly correct 'continuum' from non-tornadic to tornadic phenomena, but with the shelf-clouds intermixed among the tornadic clouds, and the MDS results further showing that clouds with upper- and mid-level tornadic features are likely to be confused with non-tornadic clouds.

Even though the precise d' and c values observed here cannot be confidently extrapolated to a real world setting, the patterns of results have implications for hazard training, creating risk communications, and anticipating inappropriate responses to environmental cues. Specifically, they show that lay observers have acquired some ability to assess tornadic potential, which could be improved were it possible to remedy their biases. The results also provide quantitative measures of behaviour for use in evaluating policy options (e.g., alternative warning strategies).

Our results suggest that training should take advantage of the attributes that emerged as most salient, namely the 'darkness' and 'clarity' of cloud scenes. It should also emphasize the difference between shelf clouds and more tornadic formations (e.g., wall clouds), while devoting extra attention to upper- and mid-level tornadic cloud features as well as mammatus clouds. Testing the possibility of such training is a logical next step in the research, which might be pursued not only with lay observers, but also with storm spotters, especially in light of the increased interest in their performance (Jans *et al* 2011, Jans and Keen 2012, Klenow and Reibestein 2014). Deficiencies in storm spotters' d 's would indicate the need for improved training to increase perceptual acuity; MDS results could suggest how to direct that training. As an anonymous reviewer remarked, forecasters would have difficulty interpreting the reports of storm spotters with biased decision criteria, especially if they vary across spotters, making the implementation of NWS policy more difficult.

Better discrimination can also reduce a natural concern of forecasters: having FA rates so high that people tune out their warnings (a 'cry wolf' effect) (Brooks 2004, Barnes *et al* 2007, Simmons and Sutter 2009, League *et al* 2010, Durage *et al* 2012, Brotzge and Donner 2013, Ripberger *et al* 2014). Since 2007, the NWS has used warning polygons in order to make its warnings more specific to the areas under threat, in contrast to the broader, county-wide warnings still practiced in Canada or the heterogeneous European warning strategies (NWS 2008, Rauhala and Schultz 2009, Durage *et al* 2012, Environment Canada 2014). Our results suggest an alternative strategy: tailor tornado warnings to the cues available to the public, based on proximity to the storm system. For example, individuals not directly under a tornadic storm might hear a warning, but only have upper- or mid-level cues

available to them. A forecast could be more useful to these individuals if it provided grounds for concern regarding features that forecasters can see, but these individuals cannot, such as the storm's movement and radar-visible signs of tornado threat (e.g., a *hook echo*) (Markowski 2002). In addition to providing people more time to prepare for storms, that additional information could also help them to judge forecasters more fairly (see also Ripberger *et al* 2014). Of course, the relevance of our results for creating interventions to improve protective action decision-making and communications depends upon the extent to which the public relies on physical environmental cues, which is likely a function of demographic variables, experience, and locale, and an empirical question that should be addressed in future work.

Conclusion

The novel application of psychophysics presented here provides a proof-of-concept of an emerging approach to risk problems: that of understanding human physical perception (Agdas *et al* 2012). These methods could be applied to cues regarding both immediate risks (e.g., flash floods, influenza, Ebola) and long-term ones (e.g., living in a tornado- or hurricane-prone region, or near a nuclear reactor). Understanding what counts as 'signal' could improve the evaluation and effectiveness of risk communications (Morss and Hayden 2010, Fischhoff 2013, Ripberger *et al* 2014), as well as our understanding of processes such as how perception of climate change is influenced by experience of extreme weather events (Leiserowitz 2006, Akerlof *et al* 2013, Broomell *et al* 2015, Lefevre *et al* 2015, van der Linden 2015).

Acknowledgments

BD, BF, SB, and AD designed the experiment. BD collected the data, and BD, BF, SB, and AD conducted the analyses. BD and BF wrote the manuscript. The research was supported by a Social Sciences and Humanities Research Council of Canada Doctoral Fellowship (awarded to BD) and by the centre for Climate and Energy Decision Making (under a cooperative agreement between Carnegie Mellon University and the National Science Foundation (SES-0949710 and SES-1463492)). The authors thank Roger Edwards and Marko Korošec for their permission to use their photographs, and Julie Demuth (National Center for Atmospheric Research), Gabrielle Wong-Parodi (Carnegie Mellon University), Michael Yu (Carnegie Mellon University), Coreen Farris (RAND Corporation), Roger Edwards, and three anonymous referees for their invaluable feedback.

References

- Agdas D, Webster G D and Masters F J 2012 Wind speed perception and risk *PLoS One* **7** e49944
- Akerlof K, Maibach E W, Fitzgerald D, Cedeno A Y and Neuman A 2013 Do people 'personally experience' global warming, and if so how, and does it matter? *Global Environ. Change* **23** 81–91
- Baird J C and Noma E J 1978 *Fundamentals of Scaling and Psychophysics* (New York: Wiley)
- Barnes L R, Grunfest E C, Hayden M H, Schultz D M and Benight C 2007 False alarms and close calls: a conceptual model of warning accuracy *Weather Forecast.* **22** 1140–7
- Blanchard-Boehm R D and Cook M J 2003 Risk communication and public education in Edmonton, Alberta, Canada on the 10th anniversary of the 'Black Friday' tornado *Int. Res. Geogr. Environ. Educ.* **13** 38–54 (<http://eric.ed.gov/?id=EJ841076>)
- Bland J M and Altman D G 1997 Cronbach's alpha *Br. Med. J.* **314** 572
- Borg I and Groenen P J F 2005 *Modern Multidimensional Scaling: Theory and Applications* (Berlin: Springer)
- Borg I, Groenen P J F and Mair P 2012 *Applied Multidimensional Scaling* (Berlin: Springer)
- Brooks H E 2004 Tornado-warning performance in the past and future: a perspective from signal detection theory *Bull. Am. Meteorol. Soc.* **85** 837–43
- Broomell S B, Budescu D V and Por H-H 2015 Personal experience with climate change predicts intentions to act *Global Environ. Change* **32** 67–73
- Brotzge J and Donner W 2013 The tornado warning process: a review of current research, challenges, and opportunities *Bull. Am. Meteorol. Soc.* **94** 1715–33
- Brotzge J and Erickson S 2010 Tornadoes without NWS warning *Weather Forecast.* **25** 159–72
- Coombs C H, Dawes R M and Tversky A 1970 *Mathematical Psychology: An Elementary Introduction* (Englewood Cliffs, NJ: Prentice-Hall)
- Cronbach L J 1951 Coefficient alpha and the internal structure of tests *Psychometrika* **16** 297–334
- Dash N and Gladwin H 2007 Evacuation decision making and behavioral responses: individual and household *Nat. Hazards Rev.* **83** 69–77
- Demuth J L 2014 personal communication
- de Zilva D and Mitchell C J 2012 Effects of exposure on discrimination of similar stimuli and on memory for their unique and common features *Q. J. Exp. Psychology* **65** 1123–38
- Doswell C A III, Moller A R and Brooks H E 1998 Storm spotting and public awareness since the first tornado forecasts of 1948 *Weather Forecast.* **14** 544–57
- Downs J S, Holbrook M B, Sheng S and Cranor L F 2010 Are your participants gaming the system? Screening Mechanical Turk workers *Proc. ACM CHI 2010 Conf. on Human Factors in Computing Systems* pp 2399–402
- Drost R 2013 Memory and decision making: determining action when the sirens sound *Weather, Clim., Soc.* **5** 43–54
- Durage S W, Wirasinghe S C and Ruwanpura J 2012 Comparison of the Canadian and US tornado detection and warning systems *Nat. Hazards* **66** 117–37
- Edwards W 1954 The theory of decision making *Psychological Bull.* **51** 380–417
- Environment Canada 2014 *Public Alerting Criteria 2014* (<http://ec.gc.ca/meteo-weather/default.asp?lang=En&n=D9553AB5-1#tornado>)
- Ericsson K A and Simon H A 1980 Verbal reports as data *Psychological Rev.* **87** 215–51
- Fagerlin A, Zikmund-Fisher B J, Ubel P A, Jankovic A, Derry H A and Smith D M 2007 Measuring numeracy without a math test: development of the subjective numeracy scale *Med. Decis. Mak.* **27** 672–80
- Fischhoff B 2013 The sciences of science communication *Proc. Natl Acad. Sci. USA* **110** (Suppl) 14033–9
- Fischhoff B and Kadwani J 2011 *Risk: A Very Short Introduction* (New York: Oxford University Press)
- Green D M and Swets J A 1988 *Signal Detection Theory and Psychophysics* (New York: Wiley)
- Groenemeijer P and Kuhne T 2014 A climatology of tornadoes in Europe: results from the European severe weather database *Mon. Weather Rev.* **142** 4775–90
- Harvey L O, Hammond K R, Lusk C M and Mross E F 1992 The application of signal detection theory to weather forecasting behavior *Mon. Weather Rev.* **120** 863–83
- Hautus M J 1995 Corrections for extreme proportions and their biasing effects on estimated values of d' *Behav. Res. Methods Instrum. Comput.* **27** 46–51
- Jans J J, Allison T J and Denzer D R 2011 The digital revolution of storm spotting: modernizations of training, tracking, and reporting *39th Conf. on Broadcast Meteorology/Conf. on Weather Warnings and Communication* (Boston, MA: American Meteorological Society)
- Jans J J and Keen C 2012 Enriching the modern day storm spotter through technology & education enhancements *7th Symp. on Policy and Socio-Economic Research* (Boston, MA: American Meteorological Society)
- Justin P, Olsson H and Bjorkman M 1997 Brunswikian and Thurstonian origins of bias in probability assessment: on the interpretation of stochastic components of judgment *J. Behav. Decis. Mak.* **10** 189–209
- Klenow D J and Reibestein J L 2014 Eyes to the sky: situating the role of storm spotters in the warning and response network *J. Homeland Secur. Emergency Manage.* **11** 437–58
- Lavis Y and Mitchell C 2006 Effects of preexposure on stimulus discrimination: an investigation of the mechanisms responsible for human perceptual learning *Q. J. Exp. Psychology* **59** 2083–101
- League C E, Diaz W, Philips B, Bass E J, Kloesel K, Grunfest E and Gessner A 2010 Emergency manager decision-making and tornado warning communication *Meteorol. Appl.* **17** 163–72
- Lefevre C E, Bruine de Bruin W, Taylor A L, Dessai S, Kovats S and Fischhoff B 2015 Heat protection behaviors and positive affect about heat during the 2013 heat wave in the United Kingdom *Soc. Sci. Med.* **128** 282–9
- Leiserowitz A 2006 Climate change risk perception and policy preferences: the role of affect, imagery, and values *Clim. Change* **77** 45–72
- Lichtenstein S, Fischhoff B and Phillips L D 1982 Calibration of probabilities: the state of the art to 1980 *Judgment Under Uncertainty: Heuristics and Biases* ed D Kahneman, P Slovic and A Tversky (Cambridge: Cambridge University Press) pp 306–34
- Lindell M K 2012 Response to environmental disasters *Handbook of Environmental and Conservation Psychology* ed S D Clayton (Oxford: Oxford University Press) pp 391–413
- Lindell M K, Lu J-C and Prater C S 2005 Household decision making and evacuation in response to hurricane Lili *Nat. Hazards Rev.* **6** 171–9
- Lindell M K and Perry R W 2012 The protective action decision model: theoretical modifications and additional evidence *Risk Anal.* **32** 616–32
- Liverman D M and Wilson J P 1981 The Mississauga train derailment and evacuation, 10–16 November 1979 *Can. Geogr./Le Géogr. Can.* **25** 365–75
- Lynn S K and Barrett L F 2014 'Utilizing' signal detection theory *Psychological Sci.* **25** 1663–73
- Macmillan N A and Creelman C D 2004 *Detection Theory: A User's Guide* (New York: Lawrence Erlbaum Associates)
- Markowski P M 2002 Hook echoes and rear-flank downdrafts: a review *Mon. Weather Rev.* **130** 852–76
- Masters J 2012 2011's billion-dollar disasters: is climate change to blame? *Weatherwise* **65** 12–9
- Morss R E and Hayden M H 2010 Storm surge and 'certain death': interviews with Texas coastal residents following hurricane Ike *Weather, Clim., Soc.* **2** 174–89
- Mundy M E, Honey R C and Dwyer D M 2007 Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing *J. Exp. Psychology: Animal Behav. Process.* **33** 124–38

- Nagele D E and Trainor J E 2012 Geographic specificity, tornadoes, and protective action *Weather, Clim., Soc.* **4** 145–55
- NOAA 2011 *Weather Spotter's Field Guide* (Silver Spring, MD: NOAA)
- NOAA 2014a US tornado climatology (accessed 4 July 2014) (<http://ncdc.noaa.gov/climate-information/extreme-events/us-tornado-climatology>)
- NOAA 2014b Advanced spotters' field guide (accessed 21 May 2014) (<http://crh.noaa.gov/Image/pah/pdf/advancedguide.pdf>)
- NOAA 2014c Basic spotters' field guide (accessed 21 May 2014) (www.crh.noaa.gov/Image/pah/pdf/basicspotterguide.pdf)
- NOAA 2015 *Annual US Killer Tornado Statistics* NOAAs National Weather Service (<http://spc.noaa.gov/climo/torn/fatalmap.php?yr=2011>)
- NWS 2008 Why storm-based warnings? (accessed 30 November 2014) (<http://nws.noaa.gov/sbwarnings/>)
- NWS 2014b Spotter training presentation (<http://nws.noaa.gov/training/wxspot.php>)
- Oppenheimer D M, Meyvis T and Davidenko N 2009 Instructional manipulation checks: detecting satisficing to increase statistical power *J. Exp. Soc. Psychology* **45** 867–72
- Poulton E C and Freeman P R 1966 Unwanted asymmetrical transfer effects with balanced experimental designs *Psychological Bull.* **66** 1–8
- Rauhala J and Schultz D M 2009 Severe thunderstorm and tornado warnings in Europe *Atmos. Res.* **93** 369–80
- Ripberger J T, Silva C L, Jenkins-Smith H C, Carlson D E, James M and Herron K G 2014 False alarms and missed events: the impact and origins of perceived inaccuracy in tornado warning systems *Risk Anal.* **35** 44–56
- Simmons K M and Sutter D 2009 False alarms, tornado warnings, and tornado casualties *Weather, Clim., Soc.* **1** 38–53
- Sorensen J H 2000 Hazard warning systems: review of 20 years of progress *Nat. Hazards Rev.* **1** 119–25
- Stewart A E 2009 Minding the Weather *Bull. Am. Meteorol. Soc.* **90** 1833–41
- Stewart A E, Lazo J K, Morss R E and Demuth J L 2012 The relationship of weather salience with the perceptions and uses of weather information in a nationwide sample of the United States *Weather, Clim., Soc.* **4** 172–89
- van der Linden S 2015 The social-psychological determinants of climate change risk perceptions: towards a comprehensive model *J. Environ. Psychology* **41** 112–24
- Wallsten T S, Fillenbaum S and Cox J A 1986 Base rate effects on the interpretations of probability and frequency expressions *J. Mem. Lang.* **25** 571–87
- Wickens T D 2001 *Elementary Signal Detection Theory* (Oxford: Oxford University Press)
- Zikmund-Fisher B J, Smith D M, Ubel P A and Fagerlin A 2007 Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitation *Med. Decision Making* **27** 663–71