

Evidence affects hypothesis judgments more if accumulated gradually than if presented instantaneously

Jennifer C. Whitman · Todd S. Woodward

© Psychonomic Society, Inc. 2011

Abstract In a hypothesis comparison task involving quantifiable evidence, we investigated whether judgments of relative probability were affected by gradual evidence accumulation and by making a series of revised ratings, rather than a single final one. Each trial of our task required participants to rate the probability that a focal hypothesis, rather than its alternative, was correct. We manipulated (1) the strength of evidence supporting the focal hypothesis, (2) the strength of evidence supporting its alternative, and (3) whether that evidence was presented in three sequential portions (gradually accumulated evidence condition) or, instead, was all presented instantaneously (control condition). In a second experiment, we also manipulated (4) the number of successive ratings made within a trial with gradually accumulated evidence. Regardless of how many ratings were made per trial, gradual evidence accumulation increased the effects of evidence strength on ratings of relative probability.

Keywords Judgment and decision making · Relative judgment

J. C. Whitman · T. S. Woodward
Department of Psychiatry, University of British Columbia,
Vancouver, BC, Canada

J. C. Whitman · T. S. Woodward
BC Mental Health and Addictions Research Institute,
Vancouver, BC, Canada

T. S. Woodward (✉)
Room A3-A116, BC Mental Health & Addictions Research
Institute – Translational Research Building, 3rd Floor,
938, W. 28th Avenue,
Vancouver, British Columbia V5Z 4H4, Canada
e-mail: Todd.S.Woodward@gmail.com

When we judge the probability of a hypothesis, the evidence for that hypothesis is often gradually accumulated, rather than becoming available instantaneously. When a given amount of evidence is presented in sequential portions (gradually accumulated), this distributed presentation may increase the extent to which that evidence is processed. That increase in processing could increase the salience or fluency of the corresponding mental representation—in turn, increasing its perceived strength. Consequently, gradual accumulation would cause evidence consistent with a given hypothesis to be seen to support it more strongly and evidence inconsistent with a given hypothesis to be seen to refute it more strongly. In the present experiments, we assessed the perceived strength of gradually accumulated versus instantaneously accumulated evidence by measuring how strongly it affected comparative judgments made between a focal hypothesis and its alternative.

While an extensive body of research has employed paradigms in which objectively quantifiable evidence is gradually accumulated (e.g., Beach, 1968; Usher & McClelland, 2001; Woodward, Munz, Le Clerc, & Lecomte, 2009), those studies confounded gradual versus instantaneous evidence accumulation with the overall strength of evidence. The purpose of the present research was to manipulate those factors independently. In order to do so in the present study, we precisely matched our gradual and instantaneous evidence conditions in terms of the total strength of evidence supporting each hypothesis. Specifically, the evidence visible on the final event of a given gradual evidence accumulation trial was equivalent, in terms of support for the focal hypothesis and support for its alternative, to the evidence visible on the corresponding instantaneous evidence control trial. Thus, if there were no bias caused by the gradual accumulation of evidence, we

would expect the final judgments made on gradual evidence accumulation trials to equal those made on the corresponding control trials.

We also wished to be able to compare subjective probability judgments with an objective mathematical norm. To that purpose, we used a paradigm with objectively quantifiable evidence strength. This was a modified version of the well-known beads-from-a-jar task (e.g., Beach, 1968; Fischhoff & Beyth-Marom, 1983; Moritz, Woodward, & Lambert, 2007; Speechley, Whitman, & Woodward, 2010; Woodward et al., 2009). In the traditional version, participants judge the most likely origin of a short sequence of red and blue beads drawn with replacement from one of two jars, each containing a different proportion of the two colors. The mathematical norm for that task involves Bayesian updating (Beach, 1968; Fischhoff & Beyth-Marom, 1983; Lopes, 1985), where the probability of hypothesis A, rather than B, given two observed data, $p(A \sim B|D1, D2)$, is

$$\frac{p(D2|A)p(A|D1)}{p(D2|A)p(A|D1) + p(D2|B)p(B|D1)}. \quad (1)$$

It should be noted that the Bayesian norm specifies the same relative probabilities regardless of whether D1 and D2 are observed sequentially or at the same time. We modified the task to involve several types of evidence, corresponding to several bead colors. Three objects, each of a different color and known to have been drawn from one of two containers, were visible throughout each trial. Each of the two containers (corresponding to the focal hypothesis and its alternative) contained objects of four different colors. Three of the four colors, referred to hereon as the *relevant colors*, corresponded to the three objects that had been drawn from one of the containers. The fourth color in the two containers was a *filler color*, used to ensure that the total number of objects in each container was the same. On gradual evidence trials, the objects within the containers became visible one color at a time. On control trials, all of the objects within each container, regardless of color, became visible instantaneously. The Bayesian model predicts that responses would be the same whether evidence was gradually accumulated or all presented instantaneously. However, as will be discussed below and is tested in two experiments in the present study, behavioral responses differ between the instantaneous and gradual evidence conditions.

In the present experiments, we expected that gradual (rather than instantaneous) presentation would increase the extent to which each item of evidence was processed, which would, in turn, increase the salience of the corresponding mental representations. As a result of that increased salience, we expected a given amount of evidence

to seem subjectively stronger when gradually accumulated than when presented instantaneously. It follows that gradual evidence more consistent with the focal hypothesis than with its alternative would be seen to support it more strongly and would lead to a greater increase in ratings of the relative probability of the focal hypothesis. It also follows that gradual evidence less consistent with the focal hypothesis than with its alternative would be seen to refute it more strongly and would lead to a greater decrease in ratings of the relative probability of the focal hypothesis.

Experiment 1

Method

Participants Sixty-eight volunteers (50 females, 18 males) with a mean age of 27.7 years ($SD = 7.9$) participated in this experiment. Participants were recruited via posters on the University of British Columbia campus and in community centers in the greater Vancouver area and also via postings on electronic bulletin boards. All participants were reimbursed \$10 per hour for their time, plus parking and transportation expenses.

Materials and procedure On each trial of the task, participants were presented with a scene depicting three lakes, two of which were upstream from the third (see Fig. 1). Throughout each trial, 3 fish were presented in the bottom lake. The participants were told that any fish appearing in the downstream lake originated in either the left-hand upstream lake or the right-hand upstream lake. Each of the upstream lakes contained 120 fish on every trial. Although this consistency was obvious due to the visual displays, the exact number of fish was not explicitly mentioned to participants. They were discouraged from explicitly counting all of the fish in each lake, since that would slow task performance too much. The positions of the fish within each upstream lake were randomized over trials, so that any two trials with identical ratios of fish of each color would not be identical in appearance.

The four fish colors used were red, gold, black, and white. The three fish visible in the downstream lake were each of a different color. We will refer to these as the *relevant colors* and to the fourth color as the *filler color*. If an upstream lake contained a high percentage of fish of the relevant colors and a low percentage of fish of the filler color, there was a high probability that it was the origin of the fish in the downstream lake. The assignment of the four fish colors to the relevant and filler colors was randomized across trials.

Half of the trials had three events (gradually accumulated evidence), with a new relevant color becoming

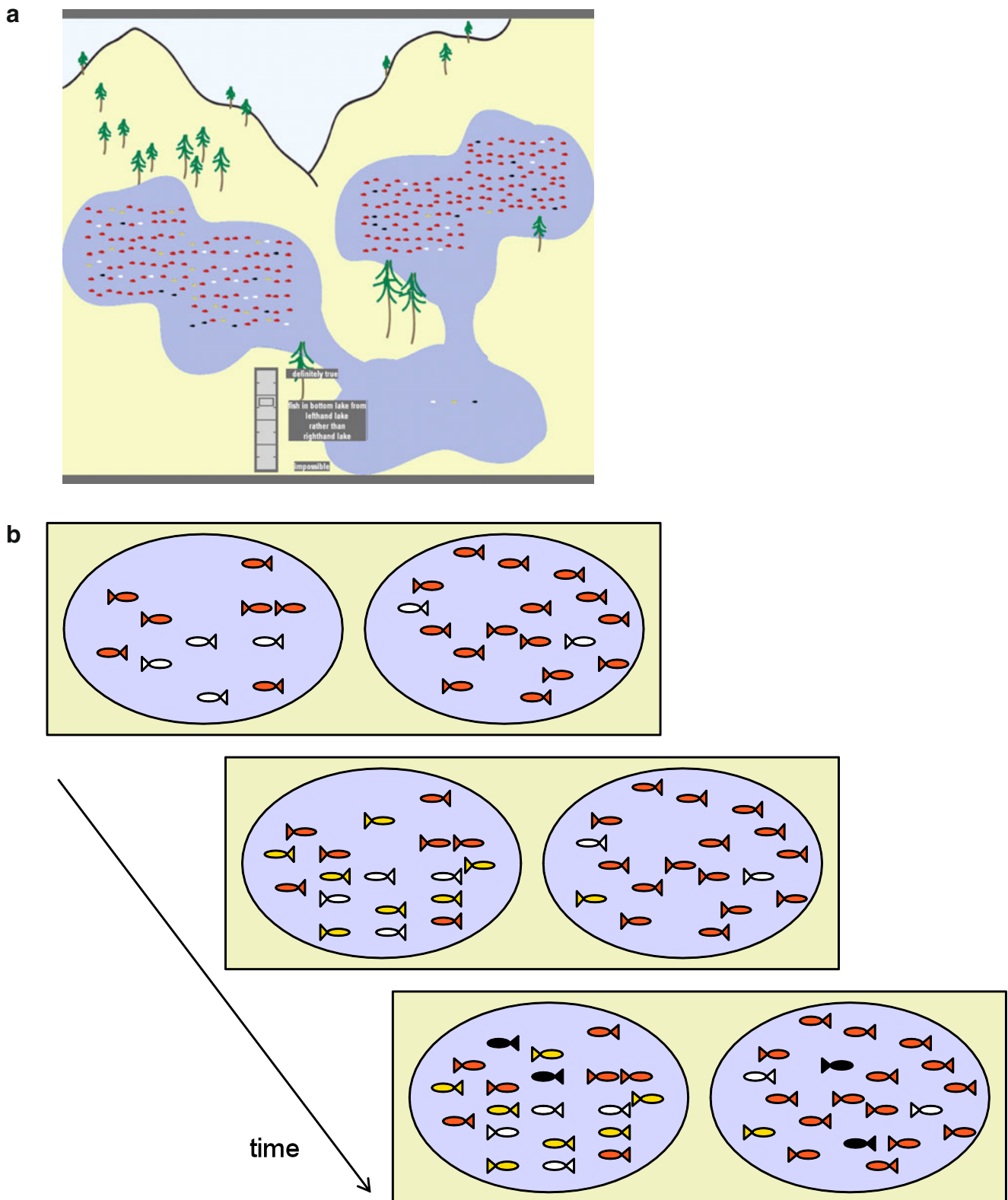


Fig. 1 a The general format of the display used throughout the experiment, with the two upstream lakes containing many fish and the rating scale next to the downstream lake. In this example, the position of the cursor on the vertical response scale represents the participant's rating of the probability that the fish in the downstream lake came from the left-hand upstream lake rather than from the right-hand one.

b A schematic representation of the sequence of events during a gradual evidence accumulation trial. In this example, the relevant colors are white, gold, and black. The filler color is red. In the upstream lakes, the white fish are visible on the first event, the gold fish become visible on the second event, and the black fish become visible on the third event

visible in the upstream lakes on each event. The other half of the trials had one event (instantaneously presented evidence). There were four levels of evidence strength for each lake, corresponding to different total numbers of fish of the relevant colors. The rest of the fish in each lake were of the filler color. Of the 120 fish in each upstream lake, the total number of fish of the relevant colors would be 24 (20%), 40 (33%), 56 (47%), or 72 (60%). As is shown in Table 1, these numbers were based on the fact that the number of fish of each individual relevant color could be either 8 or 24. Thus, there were 32 cells in the experimental design: 4 levels for the focal lake × 4 levels for the alternative lake × 2 types of evidence presentation (gradual vs. instantaneous evidence). In each of these cells, there were 4 trials, so each participant completed 128 trials in total. The assignment of the left-hand and right-hand upstream lakes to the focal and alternative hypotheses was randomized across trials.

On each event, the participant rated the probability that the three fish in the downstream lake all came from one upstream lake (the focal hypothesis) rather than from the other upstream lake (the alternative hypothesis). Ratings were made on a vertical scale, with the labels “definitely true” at the top end and “impossible” at the bottom end. To the right of the rating scale was a phrase telling participants which lake was the focal lake on the current trial (e.g., “fish in bottom lake came from left-hand lake rather than right-hand lake”). Participants used a mouse to move the slider smoothly up and down the scale and clicked the left mouse button when it was in the desired location. At the beginning of each trial, the slider was set at the midpoint of the response scale. On the second and third events of gradual evidence accumulation trials, it was set at the location corresponding to the rating made on the previous event.

Table 1 Experiment 1: Numbers of fish of the three relevant colors for each of four levels of support for a given hypothesis

Relevant Color #1	Relevant Color #2	Relevant Color #3	Total No. Fish of Relevant Colors	Total as a Percentage of the 120 Fish in the Lake
8	8	8	24	20%
8	8	24	40	33%
8	24	24	56	47%
24	24	24	72	60%

Note. In the overall experiment, there were four possible levels of support for the focal hypothesis and four possible levels of support for its alternative. The table above outlines those four levels for just one hypothesis. The numbering of the three relevant colors above is arbitrary and does not refer to presentation order. On any given trial, the three relevant colors used for the three fish in the downstream lake were randomly selected from four possibilities: red, white, gold, and black

Results and discussion

The dependent variable reported in all of the analyses here is the final rating made after all of the evidence becomes available. This is reported for each cell in the experimental design in Table 2.

Effects of gradual evidence accumulation In order to check whether the effect of evidence strength on ratings of relative probability depended on gradual versus instantaneous evidence presentation, we ran a 4 × 4 × 2 ANOVA, with factors of *support for focal* (strength of evidence supporting the focal hypothesis; four levels), *support for alternative* (strength of the evidence supporting the alternative hypothesis; four levels), and *evidence presentation* (whether the relevant evidence was gradually accumulated or all presented instantaneously). There was a significant support for focal × evidence presentation interaction, $F(3,201) = 6.54, p < .001$. This reflected stronger effects of strength of evidence supporting the focal hypothesis in the gradually accumulated evidence condition, $F(3,201) = 93.60, p < .001, \eta^2 = .58$, than in the instantaneous evidence control condition, $F(3,201) = 75.08, p < .001, \eta^2 = .53$ (see Fig. 2). There was also a

Table 2 Experiment 1: Ratings of the probability that the focal hypothesis, rather than its alternative, is correct

Strength of Evidence Supporting the Focal Hypothesis	Strength of Evidence Supporting the Alternative Hypothesis	Rating Predicted by Bayesian Norm	Rating of Relative Probability			
			Instantaneous Evidence		Gradual Evidence	
			<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
20	20	50.0	51.2	1.2	50.1	1.4
20	33	25.0	41.8	1.4	41.3	1.4
20	47	10.0	37.0	1.7	32.5	1.9
20	60	3.6	31.4	2.3	31.3	2.9
33	20	75.0	58.6	1.5	61.9	1.7
33	33	50.0	51.0	0.8	50.9	0.6
33	47	25.0	43.3	1.1	42.4	1.2
33	60	10.0	36.8	1.7	36.8	1.7
47	20	90.0	65.5	1.9	68.3	1.9
47	33	75.0	61.0	1.1	59.5	1.2
47	47	50.0	51.8	0.6	50.3	0.6
47	60	25.0	44.4	1.4	43.1	1.6
60	20	96.4	66.9	2.7	73.0	2.3
60	33	90.0	65.0	1.7	68.5	1.8
60	47	75.0	59.0	1.4	61.6	1.5
60	60	50.0	50.6	1.4	51.0	1.4

Note. Strength of evidence for each hypothesis is expressed as a percentage of the fish in the corresponding lake that are of the relevant colors. Ratings are expressed as a percentage of the height of the response scale

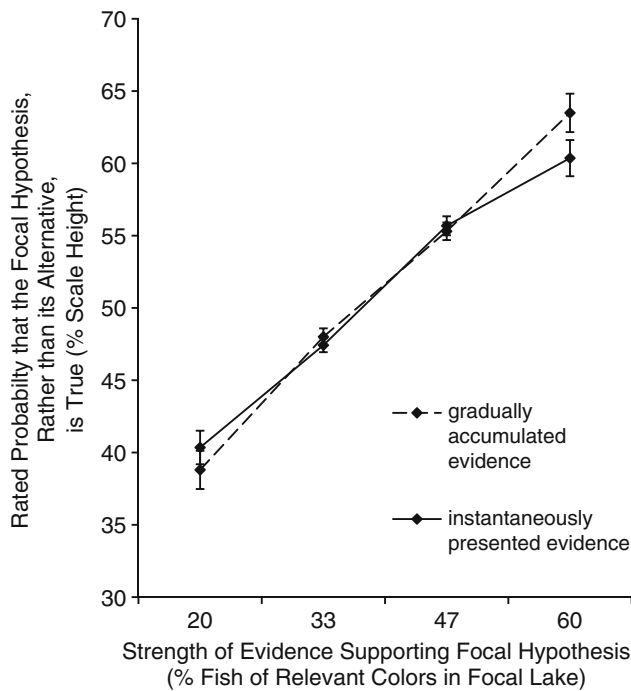


Fig. 2 Experiment 1: Effect of evidence supporting the focal hypothesis on ratings of the relative probability that the focal hypothesis, rather than its alternative, is true. Error bars represent the standard error of the mean. Varying the strength of the evidence supporting the focal hypothesis from 20% to 60% had a larger effect on ratings of relative probability (a larger increase) if evidence was gradually accumulated (dashed line) than if it was all presented instantaneously (solid line)

significant support for the alternative \times evidence presentation interaction, $F(3,201) = 4.96$, $p < .01$. This reflected the fact that ratings of the relative probability of the focal hypothesis decreased more strongly as a function of strength of evidence supporting the alternative hypothesis in the gradually accumulated evidence condition, $F(3,201) = 79.11$, $p < .001$, $\eta^2 = .54$, than in the instantaneous evidence control condition, $F(3,201) = 67.03$, $p < .001$, $\eta^2 = .50$ (see Fig. 3). Each of the two interactions reported above indicate that the gradual accumulation of evidence would cause that evidence to be weighted more heavily than in the control condition. However, in this experiment, the effects of gradual evidence accumulation were confounded with the effects of making multiple revised ratings. Experiment 2 was designed to dissociate these effects.

Experiment 2

The goals of Experiment 2 were (1) to replicate the findings of Experiment 1 and (2) to dissociate the effects of gradual evidence accumulation from the effects of making multiple revised ratings.

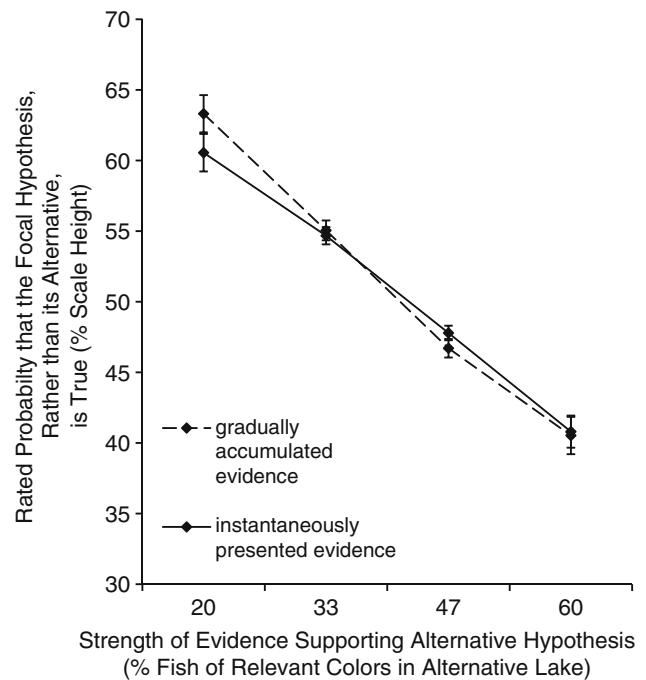


Fig. 3 Experiment 1: Effect of evidence supporting the alternative hypothesis on ratings of the relative probability that the focal hypothesis, rather than its alternative, is true. Error bars represent the standard error of the mean. Varying the strength of the evidence supporting the alternative hypothesis (from 20% to 60%) had a larger effect on ratings of relative probability (a larger decrease) if evidence was gradually accumulated (dashed line) than if it was all presented instantaneously (solid line)

Method

The method in Experiment 2 was the same as that in Experiment 1, with the following exceptions.

Participants Sixty-five volunteers (40 females, 25 males) with a mean age of 25.5 years ($SD = 7.0$) participated in this experiment.

Materials and procedure In order to minimize the overall duration of the experiment, we reduced the number of levels of evidence strength for each hypothesis from four to three. Once all of the evidence was visible, each lake contained 20%, 40%, or 60% fish of the relevant colors. The numbers of fish corresponding to each level of evidence strength are outlined in Table 3. There were 36 cells in the experimental design: 3 levels for the focal lake \times 3 levels for the alternative lake \times 4 types of evidence presentation (three types of gradual evidence and one type of instantaneous evidence). In each of these cells, there were 4 trials, so each participant completed 144 trials in total.

We changed the sequence of events within a gradual evidence trial as follows. The display alternated between

Table 3 Experiment 2: Numbers of fish of the three relevant colors for each of three levels of support for a given hypothesis

Relevant Color #1	Relevant Color #2	Relevant Color #3	Total No. Fish of Relevant Colors	Total as a Percentage of the 120 Fish in the Lake
4	16	4	24	20%
4	16	28	48	40%
28	16	28	72	60%

Note. In the overall experiment, there were three possible levels of support for the focal hypothesis and three possible levels of support for its alternative. The table above outlines those three levels for just one hypothesis. The numbering of the three relevant colors above is arbitrary and does not refer to presentation order. On any given trial, the three relevant colors used for the three fish in the downstream lake were randomly selected from four possibilities: red, white, gold, and black

periods when the fish in the upstream lakes were visible and periods in which the cursor on the rating scale could be moved. After viewing the fish in the upstream lakes, participants clicked the left mouse button to make the rating scale cursor movable, which also made the upstream fish disappear. On some gradual evidence trials, this caused the rating scale cursor to become movable on all three events. On others, it caused it to be movable only on the second and third events or only on the third event. Participants never knew ahead of time what type of gradual evidence trial they were on and, thus, could not predict on any given event whether or not they would be required to make a rating after clicking the left mouse button. This ensured that participants always attended to the contents of the upstream lakes before clicking the mouse to advance through the trial. Otherwise, on gradual evidence trials requiring ratings only on the last event, participants might have simply clicked the mouse quickly to advance to the

end of the trial without even looking at the contents of the upstream lakes. Such a strategy would have effectively changed a gradual evidence trial into an instantaneous evidence trial. Instead, we designed the experiment so that participants were required to pay attention to the contents of the upstream lakes on each event and to prepare an accurate response in case they were required to make a rating.

Since the instructions for Experiment 2 were somewhat more complex than those for Experiment 1, we started each testing session with a practice session. This involved four blocks of four trials each, with verbal instructions given before each block. Each block contained only one of the four response conditions. The first was the instantaneous evidence condition, which was followed by the gradual evidence condition with a rating made on each event, then the gradual evidence condition with a rating made on the last two events, and then the gradual evidence condition with a rating made on the last event only. Once participants completed this practice session, they were informed that the four conditions would be randomly intermixed in the main experiment.

Results and discussion

The dependent variable reported in all of the analyses here is the final rating made after all of the evidence becomes available. This is reported for each cell in the experimental design in Table 4.

Replication of experiment 1 In order to test for replication of the results reported in Experiment 1, we submitted the data from the gradual evidence condition with ratings on each event and the data from the instantaneous evidence condition to a 3 × 3 × 2 ANOVA. This had factors of *support for focal* (strength of evidence supporting the focal hypothesis; three

Table 4 Experiment 2: Ratings of the probability that the focal hypothesis, rather than its alternative, is correct

Strength of Evidence Supporting the Focal Hypothesis	Strength of Evidence Supporting the Alternative Hypothesis	Rating Predicted by Bayesian Norm	Rating of Relative Probability							
			Instantaneous Evidence		Gradual Evidence, Rate on All Three Events		Gradual Evidence, Rate on Second and Third Events		Gradual Evidence, Rate on Third Event Only	
			M	SE	M	SE	M	SE	M	SE
20	20	50.0	49.6	0.4	49.8	0.4	49.5	0.5	49.8	0.5
20	40	12.5	33.3	1.1	29.4	1.1	29.0	1.1	29.0	1.1
20	60	2.0	26.9	1.2	18.9	1.2	20.2	1.2	21.4	1.2
40	20	87.5	67.1	1.3	71.9	1.4	71.8	1.2	70.1	1.6
40	40	50.0	49.7	0.4	49.8	0.4	49.7	0.5	49.9	0.4
40	60	12.5	31.1	1.1	28.8	1.2	31.7	1.0	30.8	1.1
60	20	98.0	75.4	1.4	81.6	1.4	78.1	1.8	79.5	1.2
60	40	87.5	67.7	1.2	71.9	1.2	69.4	1.2	71.5	1.2
60	60	50.0	49.8	0.4	50.3	0.5	49.7	0.6	50.1	0.5

Note. Strength of evidence for each hypothesis is expressed as a percentage the fish in the corresponding lake that are of the relevant colors. Ratings are expressed as a percentage of the height of the response scale

levels), *support for alternative* (strength of the evidence supporting the alternative hypothesis; three levels), and *evidence presentation* (whether the relevant evidence was gradually accumulated or all presented instantaneously). As in Experiment 1, there was a significant support for focal \times evidence presentation interaction, $F(2,128) = 51.07$, $p < .001$. This reflected the fact that ratings of the relative probability of the focal hypothesis increased more strongly as a function of strength of evidence supporting the focal hypothesis in the gradually accumulated evidence condition, $F(2,128) = 616.06$, $p < .001$, $\eta^2 = .91$, than in the instantaneous evidence control condition, $F(2,128) = 394.93$, $p < .001$, $\eta^2 = .86$ (see Fig. 4). There was also, as there was in Experiment 1, a significant support for the alternative \times evidence presentation interaction, $F(2,128) = 35.74$, $p < .001$. This reflected the fact that ratings of the relative probability of the focal hypothesis decreased more strongly as a function of strength of evidence supporting the alternative hypothesis in the gradually accumulated evidence condition, $F(2,128) = 567.10$, $p < .001$, $\eta^2 = .90$, than in the instantaneous evidence control condition, $F(2,128) = 400.88$, $p < .001$, $\eta^2 = .86$ (see Fig. 5).

Effects of gradual evidence accumulation controlling for number of ratings In Experiment 2, we included a condition in which evidence was gradually accumulated but only a single rating was made, rather than a rating after each of the three portions of evidence, as in the analysis above. This allowed us to test for an effect of gradual evidence accumulation when the number of ratings made was held constant. This involved a $3 \times 3 \times 2$ ANOVA, with factors of *support for focal* (three levels), *support for alternative* (three levels), and *evidence presentation* (gradual vs. instantaneous).

As in all the analyses presented above for Experiments 1 and 2, we found a significant support for the focal \times evidence

presentation interaction, $F(2,128) = 29.41$, $p < .001$. This reflected the fact that ratings of the relative probability of the focal hypothesis increased more strongly as a function of support for the focal hypothesis in the gradually accumulated evidence condition, $F(2,128) = 490.47$, $p < .001$, $\eta^2 = .89$, than in the instantaneous evidence control condition, $F(2,128) = 394.93$, $p < .001$, $\eta^2 = .86$ (see Fig. 4). Second, as in all the analyses presented above for Experiments 1 and 2, we found a significant support for the alternative \times evidence presentation interaction, $F(2,128) = 17.31$, $p < .001$. This reflected the fact that ratings of the relative probability of the focal hypothesis decreased more strongly as a function of strength of the evidence supporting the alternative hypothesis in the gradually accumulated evidence condition, $F(2,128) = 548.96$, $p < .001$, $\eta^2 = .90$, than in the instantaneous evidence control condition, $F(2,128) = 400.88$, $p < .001$, $\eta^2 = .86$ (see Fig. 5). This analysis confirms that evidence affects ratings of relative probability more when gradually accumulated than when presented instantaneously, even if the total number of ratings made is controlled for. Thus, we can conclude that the effect is due to gradual evidence accumulation, rather than to the act of making successive ratings.

General discussion

The main goal of the present research was to determine how gradual evidence accumulation affects comparative judgments. Those judgments were assessed via ratings of the relative probability that a focal hypothesis, rather than its alternative, was true. In two experiments, we found a given amount of evidence to affect ratings of relative probability more if gradually accumulated (presented in three sequential portions) than if presented instantaneously. Our investigation

Fig. 4 Experiment 2: Effect of evidence supporting the focal hypothesis on ratings of the relative probability that the focal hypothesis, rather than its alternative, is true. Error bars represent the standard error of the mean. Varying the strength of the evidence supporting the focal hypothesis (from 20% to 60%) had a larger effect on ratings of relative probability (a larger increase) if evidence was gradually accumulated (dashed and dotted lines) than if it was all presented instantaneously (solid line)

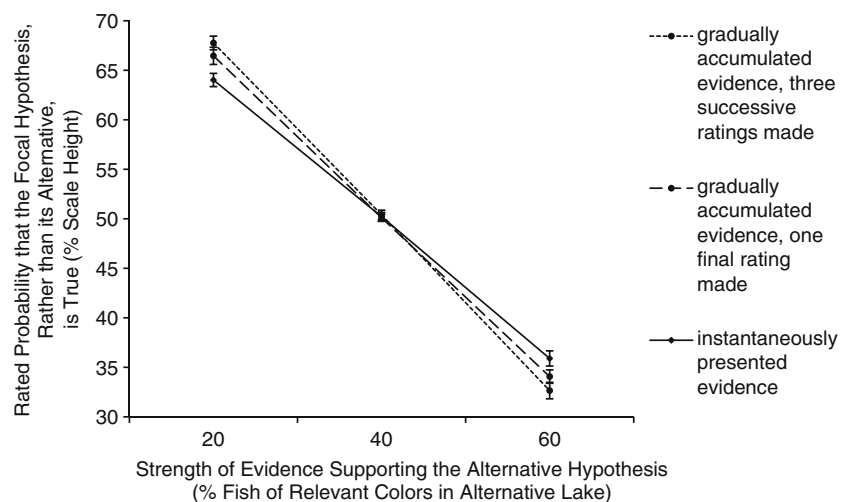
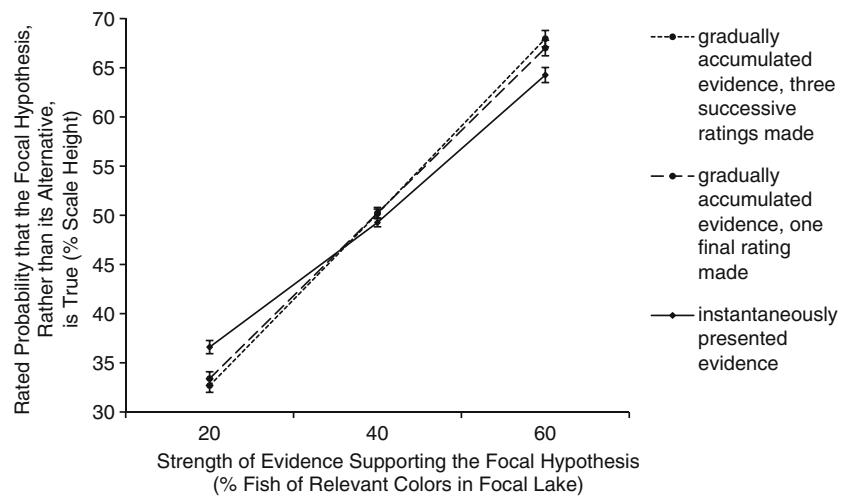


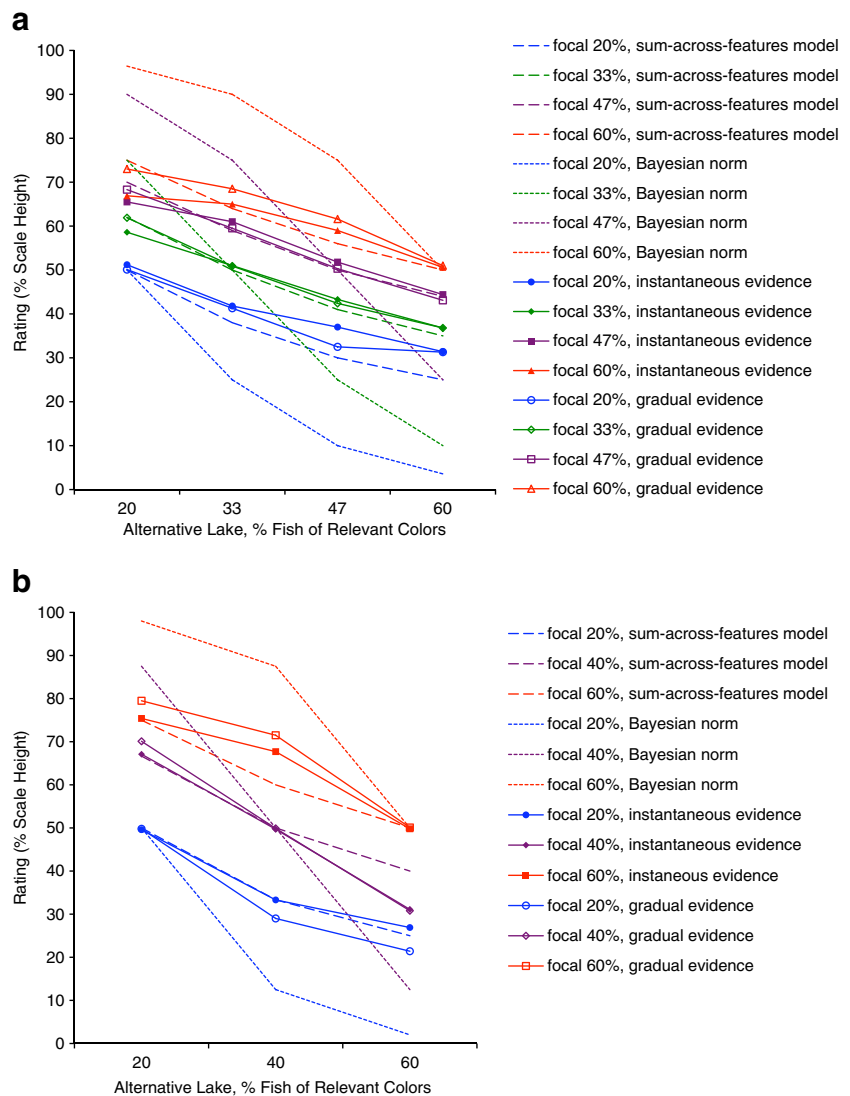
Fig. 5 Experiment 2: Effect of evidence supporting the alternative hypothesis on ratings of the relative probability that the focal hypothesis, rather than its alternative, is true. Error bars represent the standard error of the mean. Varying the strength of the evidence supporting the alternative hypothesis (from 20% to 60%) had a larger effect on ratings of relative probability (a larger decrease) if evidence was gradually accumulated (dashed and dotted lines) than if it was all presented instantaneously (solid line)



diverged from previous studies with objectively quantifiable evidence in that we kept the strength of evidence constant while manipulating whether it was accumulated gradually or

all presented instantaneously. Thus, the key contribution of this work is that it dissociates the effects of gradual evidence accumulation from the effects of changing evidence strength.

Fig. 6 Comparison of actual ratings made to those predicted by the Bayesian norm and the sum-across-features model. Actual ratings made (solid lines) are clearly closer to those predicted by the sum-across-features model than by the Bayesian norm. **a** Experiment 1 data. **b** Experiment 2 data: The gradual evidence trials for which the results are depicted here are those on which a rating was made only on the last of three events



While some previous experiments (Hogarth & Einhorn, 1992) manipulated whether evidence was gradually accumulated versus presented instantaneously, they produced inconsistent results, reporting an effect of gradual versus instantaneous evidence in some experiments, but not in others. Given that evidence strength in those experiments could not be objectively quantified and that the gradual versus instantaneous manipulation was between subjects, their gradual versus instantaneous manipulation was necessarily confounded with individual differences in subjective evidence strength. In addition, that work did not manipulate gradual evidence presentation independently from making multiple successive ratings. Thus, our work is the first to report a replicable effect whereby gradual accumulation of objectively quantifiable evidence increases subjective strength. It is also the first to show that this effect is not due to making successively revised ratings.

A secondary finding evident in our data is that participants' responses were conservative relative to the Bayesian norm, as has been reported previously in the beads task (Anderson, 1981; Juslin, Nilsson, & Winman, 2009; Lopes, 1985; Shanteau, 1975). Such conservatism has been attributed to misaggregation of evidence over time, with the combined value of several sequential samples being a weighted average of their individual scale values (Anderson, 1981; Lopes, 1985; Shanteau, 1975). If conservatism resulted from such misaggregation, one would predict more conservatism in response to gradually accumulated evidence than in response to the same evidence presented instantaneously. Interestingly, our results showed the opposite, indicating the need for further theoretical accounts.

In our data, the effect of gradual evidence accumulation was present regardless of whether the evidence favored the focal hypothesis or favored its alternative. If gradually accumulated, evidence supporting the focal hypothesis was seen to support it more strongly, while evidence supporting the alternative was seen to refute the focal hypothesis more strongly. These results are consistent with our prediction that processing evidence to a greater extent would lead to more salient mental representations and, consequently, greater subjective strength. This tendency to give increased weight to evidence divided into three gradually accumulated portions can be thought of as a new type of packing effect. This is analogous to the effect observed with packed versus unpacked hypotheses (e.g., Bonini & Gonzalez, 2005; Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). A packed hypothesis (e.g., that someone will die of natural causes) is rated as less probable than its unpacked alternative (e.g., that they will die of cancer, a heart attack, a stroke, or some other natural cause). One explanation for the tendency to rate unpacked hypotheses as more

probable than packed ones is that explicitly mentioning one of the unpacked components of a hypothesis increases its salience, thereby increasing its perceived support (Rottenstreich & Tversky, 1997). Similarly, we found in the present experiments that unpacked evidence (evidence presented in three portions) seemed to be subjectively more salient than packed evidence (the same evidence presented in a single portion), possibly as a result of being processed more extensively.

A secondary finding, evident in Tables 2 and 4, is that participants' responses deviated substantially from those predicted by the Bayesian norm. An alternative model, which we will call the *sum-across-features* model, would involve participants calculating the total number of fish of the relevant colors in each lake (summing across three colors). They would then compare the number of fish of relevant colors in the focal lake with the total number of fish of relevant colors summed across both upstream lakes, as in Luce's choice axiom (Luce, 1977). Figure 6 demonstrates that participants' responses are much more closely aligned to this sum-across-features model than to the Bayesian one. However, the Bayesian model and participants' responses both follow a bow-shaped pattern, while the sum-across-features model does not. Neither model accounts for the difference in ratings between the gradual and instantaneous evidence conditions. The development of a cognitive model that does account for this effect remains a direction for future research.

Another interesting direction for future research would be a parallel investigation requiring participants to remember evidence in long-term memory. Such an investigation would allow application of these results to a wider range of everyday situations. For example, a political party's campaign could involve revealing all of the points in their platform at the same time or else gradually revealing them over a period of days or weeks. On the basis of the results reported here, one would predict that revealing each point of a political platform in sequence would elicit more political support than would presenting all of those points instantaneously.

Acknowledgments The authors thank Patrick Carolan, Jessica Ferguson-King, Alissa Low, and Emily Stewart for assistance with data collection and data management.

Funding This research was funded by a Natural Sciences and Engineering Research Council of Canada Operating Grant (RGPIN 343016–07), a Canadian Institutes of Health Research New Investigator Award (T.S.W.), a Michael Smith Foundation for Health Research Scholar Award (T.S.W.), a Natural Sciences and Engineering Research Council of Canada Post Graduate Scholarship Award (J.C.W.), and a Michael Smith Foundation for Health Research Senior Trainee Award (J.C.W.).

References

- Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Beach, L. R. (1968). Probability magnitudes and conservative revision of subjective probabilities. *Journal of Experimental Psychology*, *77*, 57–63.
- Bonini, N., & Gonzalez, M. (2005). Inconsistent probability estimates of a hypothesis. *Experimental Psychology*, *52*, 55–66.
- Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*, 239–260.
- Hogarth, R. M., & Einhorn, H. J. (1992). Ordereffects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.
- Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review*, *116*, 856–874.
- Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian-inference. *Bulletin of the Psychonomic Society*, *23*, 509–512.
- Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*, 215–233.
- Moritz, S., Woodward, T. S., & Lambert, M. (2007). Under what circumstances to patients with schizophrenia jump to conclusions? A liberal acceptance account. *British Journal of Clinical Psychology*, *46*, 127–137.
- Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review*, *104*, 406–415.
- Shanteau, J. C. (1975). Averaging versus multiplying combination rules of inference judgment. *Acta Psychologica*, *39*, 83–89.
- Speechley, W. J., Whitman, J. C., & Woodward, T. S. (2010). The contribution of hypersalience to the “jumping to conclusions” bias associated with delusions in schizophrenia. *Journal of Psychiatry & Neuroscience*, *35*, 7–17.
- Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*, 467–547.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
- Woodward, T. S., Munz, M., LeClerc, C., & Lecomte, T. (2009). Change in delusions is associated with change in “jumping to conclusions”. *Psychiatry Research*, *170*, 124–127.