Received: 10 December 2021 | Revised: 11 August 2022 | Accepted: 12 August 2022

DOI: 10.1002/bdm.2298

RESEARCH ARTICLE

WILEY

Combining verbal forecasts: The role of directionality and the reinforcement effect

Karl Halvor Teigen[1] | Marie Juanchich[2] | Erik Løhre[3]

[1]University of Oslo, Oslo, Norway

[2]University of Essex, Colchester, UK

[3]BI Norwegian Business School, Oslo, Norway

Correspondence
Karl Halvor Teigen, Department of Psychology, University of Oslo, Oslo, Norway.
Email: k.h.teigen@psykologi.uio.no

Abstract

Recent research has shown that when people combine verbal probabilistic forecasts from two sources, they are not simply averaged but can reinforce each other; so when two advisors both said an event was "rather likely," some listeners concluded that the event was "quite likely". Conversely, when both said the event was "rather unlikely," people concluded that it was "quite unlikely." The present studies demonstrate that the direction of this effect is not evoked by high versus low probabilities, but by the *directionality* of verbal probability expressions. Some phrases are affirmative, directed towards occurrences ("there is a chance"), whereas others are negations, pointing to the possibility that the event might not occur ("it is not certain"). Two positive phrases are perceived to reinforce each other, even when they convey low probabilities, resulting in a higher combined probability estimate, whereas two negative phrases do the opposite, regardless of the probabilities they convey. We show that this effect occurs both for equal and unequal verbal phrases, regardless of the probability equivalents of the expressions. We also found a positive, but weaker, reinforcement effect of numerical probabilities.

KEYWORDS
combining forecasts, directionality, uncertainty, verbal probabilities

1 | INTRODUCTION

Experts in many fields, from finance to medicine to sports, are often called upon to make informed predictions about future events. Will stocks be rising during the year to come? Will the Norwegian team qualify for the next World Cup in football? When will we see the end of the pandemic? Such forecasts can rarely be issued with certainty and must accordingly be qualified with verbal or numeric hedges ("it is quite likely" and "there is a 60% to 70% probability"). Much research has been devoted to comparing these two ways of expressing expectations. In an ideal world, we might think that linguistic uncertainty phrases, also known as "verbal probability expressions" (VPEs) should correspond to specific sections of the 0% to 100% numeric probability scale. To this end, conversion tables have been developed by several organizations to coordinate the use of linguistic and numeric probabilities, for instance, translating *likely* into $p > 60\%$ and vice versa, to facilitate communication (European Food Safety Authority et al., 2018; Mandel & Irwin, 2021; Mastrandrea et al., 2010). Such guidelines are not without their own limitations, as they only define a subset of all common verbal phrases, and do not capture all aspects of how words (or numbers) are used in everyday language. It turns

out that both modes do more than just reflect the strength or level of an expectation, they also convey a format-dependent message. For instance, numbers suggest objectivity and exactness (Gurmankin et al., 2004), whereas words are more perspective dependent and elastic and may reveal, through choice of terms, the speaker's attitudes and intentions to a greater extent than numbers do (Smith & Jucker, 2014). It follows that the inferences based on words are not always identical to inferences based on numbers (Collins & Hahn, 2018; Teigen & Brun, 1999).

We are in this research concerned with the combined effect of two probabilistic statements. Such situations can arise when people assess the joint occurrence of two uncertain events (conjunctions) or when forecasts of the same event are obtained from two separate sources. The problem of assessing conjunctions has been investigated extensively in the past (Fisk, 2022; Tversky & Kahneman, 1983), whereas little is known about how people combine multiple probability estimates of the same event. A recent study by Mislavsky and Gaertig (2021) demonstrated an intriguing format difference when people receive two verbal or two numerical forecasts of the same event. When two advisors both stated that hypothetical LPT stocks have a 60% to 69% probability of going up, most listeners concluded that 60% to 69% was indeed the most appropriate forecast. In this case, people did not "count" the estimates as adding up to anything beyond the stated values but saw them presumably as a replication of each other. However, when two advisors used verbal phrases instead of numbers, and both stated that it was "rather likely" that the stocks would rise, many believed that a rise was *more* likely—it could be "quite likely" or even "very likely." Conversely, two low numerical probability forecasts of 30% to 39% made listeners conclude that 30% to 39% was a proper estimate, whereas two linguistic forecasts of "rather unlikely" induced a downward shift in probability judgments, towards "quite unlikely." Mislavsky and Gaertig concluded that combining two verbal probabilities led to more extreme probability estimates than combining numbers.

2 | WHEN DO VERBAL EXPRESSIONS REINFORCE EACH OTHER?

Mislavsky and Gaertig (2021) considered the direction of the shifts towards lower or higher extremes to be a function of the level of the probabilities conveyed: Combining forecasts below 50% would lead to even lower estimates, whereas combining forecasts higher than 50% led to more extreme high estimates. However, they based their conclusions on a limited selection of verbal phrases, namely, *likely* and *unlikely* with a modifier (e.g., *rather likely*, *quite likely*, *rather unlikely* and *quite unlikely*). These phrases do not differ merely by their position on the probability dimension. They also differ in directionality (Teigen, 1988; Teigen & Brun, 1995). Some VPEs have a focus on occurrences, drawing attention to the chances of a target outcome T to occur ("T is likely" and "T is possible") implying that $p > 0$. Other VPEs are phrased as negations, highlighting the complementary outcome, namely, that T might *not* occur ("T is unlikely" and "T is not certain"; $p < 1$). Directionality should not be confused with level of probability, as even low probabilities can be expressed in a positive way (e.g., "there is a small chance") and high probabilities are occasionally expressed (framed) as a negation (e.g., "not quite sure") (Honda & Yamagishi, 2017; Juanchich et al., 2010; Piercey, 2009). In Mislavsky and Gaertig's studies, level of probability and direction were confounded. "Rather likely" indicates a high probability that is framed positively, and "rather unlikely" indicates a low probability framed as a negation.

We argue in the present paper that the reported shifts towards more extreme estimates are not caused by the location of these phrases on the probability dimension but are primarily due to their positive ($p > 0$) or negative ($p < 1$) directionalities. We propose that it is this positive or negative directionality that drives the shift when people combine verbal estimates of probabilities. When two *positive* statements about a target outcome are combined, they reinforce each other in such a way that the outcome will appear more likely, whereas two *negative* statements make the target appear more uncertain and doubtful than either statement judged in isolation. We call in the present paper the mutual strengthening of VPEs with the same directionality a *reinforcement effect*.

Mislavsky and Gaertig (2021) (hereafter M&G) investigated several potential differences between numeric and verbal estimates that could be responsible for the polarization effect, including information source (whether advisors are believed to have access to same or different information) and objectivity (whether advisors' estimates are based on opinions vs. facts), without finding strong evidence for any specific explanation. They also considered directionality as a potential mechanism, which they aimed to test in their Study 6. However, this test was performed in an indirect and, in our view, inadequate way by keeping the verbal statements unchanged, while making the numeric estimates more evaluable, by adding information about which p values should be regarded as good or bad. This was done by simply telling participants that a prediction above 50% should be regarded as a "good sign." However, this manipulation does not test the hypothesis that directionality drives the reinforcement effect, because directionality does not refer to the goodness (valence) of a prediction, but to the affirmation versus negation of a target outcome. For example, one can use a positive phrase to describe an aversive outcome (there is *a chance* that you will lose) and the other way around (I am *not sure* you will be successful). Directionality is moreover an inherent property of the verbal phrases, not of the outcomes, so the direction of a shift should be tested more directly by investigating the combination of two positive or two negative verbal phrases regardless of their location as high or low on the probability scale. It has been shown that positive and negative phrases are read as conveying opposite recommendations even when they correspond to the same probability level. A patient who is considering a "possible" cure is advised to give it a try, as opposed to when it is "uncertain" that it will be helpful (Teigen & Brun, 1999, Experiment 1). Especially relevant to the idea of shift in probability perception, phrases with opposite

directionality will be used to describe opposing trends. For example, a 30% chance of rain was more often described as "unlikely" when the chance had decreased from 50% to 30% than when it increased from 10% to 30%. In the latter case, it was better characterized with a directionally positive term (e.g., "a small chance of rain"; Juanchich et al., 2010).

3 | THE PRESENT STUDIES

The present studies were designed to test how verbal probabilities reinforce each other when combined, and whether they differ from numerical probabilities in this respect. For the combination of two probability statements of the same event (as studied by M&G), we expected directionality to be a more important factor than level of probability. We predicted that positive directionality phrases—whatever probability they conveyed—would lead to an upward adjustment (reinforcement), whereas negative verbal probabilities would more often be adjusted downwards. This means that combined estimates for the same event may become less conservative (Armstrong et al., 2015), and hence potentially less correct, as they move beyond the estimates given by the experts.

3.1 | Study 1: One versus two equal probabilities

We tested our hypotheses regarding the combination of probabilities as a function of directionality by examining the joint effect of probabilistic forecasts from two financial advisors, reproducing the original design from M&G's Study 2. In this study, participants rated the likelihood of stock increase after being exposed to one—verbal or numerical—forecast from a single expert, and then again after receiving the same estimate from a second forecaster. Would two identical forecasts make stock increase appear more likely, less likely, or equally likely than only one?

3.2 | Study 2: One versus two unequal probabilities

All M&G's studies explored the effect of two advisors uttering identical forecasts. But two advisors often diverge in their predictions. To extend the findings of Study 1 to unequal probabilities, a second study was performed where the advisors made divergent estimates of the probability of stock increase, one being more (or less) optimistic than the other.

3.3 | Study 3: Separate versus combined judgments

In Studies 1 and 2, comparisons between one versus two forecasts were done in a within-Ss design, where the same participants estimated likelihoods of stock increase based on estimates from one versus two advisors. The third study used a between-Ss design where ratings of single versus two advisors were performed by different participants.

3.4 | Study 4: Combining equal verbal probabilities

In this study, participants were asked to choose a verbal expression that described the combination of two equal positive forecasts (a chance) or two equal negative forecasts (not certain).

All studies except the last included three conditions, with advisors giving numerical, verbal positive or verbal negative probability estimates, respectively. In addition to the verbal pair *rather likely/rather unlikely* used in M&G's original studies, we included a novel pair, *a chance/not completely certain*, which also differ in directionality. The likelihood estimates were given on 1–10 rating scales in the Studies 1 and 2 and on a 0% to 100% probability scale in Study 3. In Study 4, participants chose the most appropriate verbal expression.

The hypotheses, material and analytical plans for the four studies were preregistered prior to data collection of the studies. The preregistrations are available on the Open Science Framework here: https://osf.io/qnm4f/?view_only=533b57ff2e9445a88f62b02941441f8b. The plans originally included a set of studies of conjunctive probabilities of positive and negative terms, using the same selection of numeric and verbal terms. These studies turned out to be difficult to integrate—theoretically and thematically—with the combination studies and were for reasons of simplicity omitted from the present manuscript and will be presented and discussed in a separate report.

4 | STUDY 1: ONE VERSUS TWO EQUAL FORECASTS

This study examined combinations of two equal numerical or verbal forecasts of the same event. The purpose was twofold: first, to replicate the findings of Mislavsky and Gaertig (2021), using a stock scenario with the same numerical estimates and the same verbal phrases as in their original studies, with a neutral rather than a labeled rating scale; second, to investigate the effects of two novel verbal phrases, *a chance* and *not completely certain*.

4.1 | Method

4.1.1 | Participants

Participants were recruited online from the United Kingdom and Ireland via the crowdsourcing platform Prolific. After excluding participants who failed a simple attention check or spent less than 1 min on the whole survey, the final sample consisted of 597 participants (419 women, 174 men, and 4 other), with ages ranging from 18 to

66 years ($M = 32.0$, $SD = 9.9$). About half of them (50.9%) were full-time employed, and 43.2% reported to have a bachelor's degree or equivalent.

4.1.2 | Design

Participants were randomly assigned to six experimental conditions: two numerical (low and high) and four verbal conditions, corresponding to the crossing of two independent variables with two levels each: 2 probability level (low vs. high) × 2 directionality (positive vs. negative). Two of the resulting four conditions (low negative and high positive) were the same as used by M&G in their Study 2, whereas two (low positive and high negative) were new, designed to disentangle the effect of the probability from that of the directionality.

4.1.3 | Questionnaires

Participants in the two numerical conditions received the following stock price scenario, copied from M&G.

"The company Liberty Property Trust (LPT) has a present stock price of $39.90. How likely is it that LPT stocks will increase their value and close above $39.90 in a year from now? Two analysts, A and B, give their predictions, as described below.

Analyst A: "It is 30–40% [60–70%] likely that LPT will close above $39.90 in a year from now."

Based on this single analyst's prediction, how likely do you think it is that LPT stocks will close above $39.90 in a year from now?

Analyst B: "It is 30–40% [60–70%] likely that LPT will close above $39.90 in a year from now."

Based on these two analysts' predictions, how likely do you think it is that LPT stocks will close above $39.90 in a year from now?"

In two verbal conditions the analysts said an increase was "rather likely" (positive, high probability level) or "rather unlikely" (negative, low probability), as in the original study. In two novel verbal conditions, the analysts said there was "a chance" (positive, low) or that it was "not completely certain" (negative, high). "A chance" has in previous studies been shown to correspond to probabilities below 50% (O'Brien, 1989) and "not completely certain" to probabilities above 50% (Juanchich et al., 2019).

Likelihood judgments of both single and combined forecasts were performed on 10-point scales from 1: *nearly impossible* to 10: *nearly certain*, with only the endpoints labeled. In contrast, in M&G's original studies, all steps on the scales were labeled, either with numerical probabilities (e.g., 1 = 0% to 9% and 2 = 10% to 19%) or with verbal labels (e.g., 1 = nearly impossible and 2 = extremely unlikely). The scale used in the present study gave less guidance regarding what a specific scale value was supposed to mean. However, it ensured that participants in all conditions were responding on the same scale.

4.1.4 | Manipulation Check 1: Awareness of communicative intent

Probabilistic statements indicate, by definition, that the results can turn out in more than one way. A probability of 30% to 40% implies both that the stocks might go up or that they might not go up. But recipients of this message might favor one of these interpretations above the other, based on the level of probability and the expression's directionality. As a control of the directionality, participants were asked:

When someone says: "It is 30–40% likely [60–70% likely] [rather unlikely] [rather likely] [a chance] [not completely certain] that the stocks will go up," do they try to direct your attention to the possibility that …

a. the stocks might go up
b. the stocks might not go up
c. both.

4.1.5 | Manipulation Check 2: Numeric equivalents

As a final control of the numeric meaning of the VPEs, participants were assigned randomly to one of four "translation" conditions and asked:

- What is the probability that an event occurs when described as being "rather likely" ["rather unlikely"] [having "a chance"] [being "not completely certain"] to occur?

  About … % (fill in a number between 0 and 100).

4.2 | Results

4.2.1 | Manipulation checks

The verbal phrases "rather unlikely" and "a chance" were assumed to be instances of "low" probabilities ($p < 50\%$), whereas "rather likely" and "not completely certain" were assumed to represent "higher" probabilities ($p > 50\%$). This was confirmed by participants' own judgments of these terms, when they assessed them on their own and without context. "Rather unlikely" conveyed a probability of $M = 26.7\%$ ($SD = 18.5$) and "a chance" a probability of $M = 41.7\%$ ($SD = 15.7$). In contrast, "rather likely" received a probability equivalent of $M = 67.7\%$ ($SD = 12.6$) and "not completely certain" $M = 52.1\%$ ($SD = 20.6$). The probability equivalents for the novel pair of VPEs were closer to 50%, yet the positive phrase was still significantly *lower* than the negative, $t(302) = 4.97$, $p < .001$, Cohen's $d = 0.57$.

Awareness of directionality was checked by the question about how statements directed listeners' attention. As shown in the left panel of Figure 1, numerical probabilities (both low and high) were not

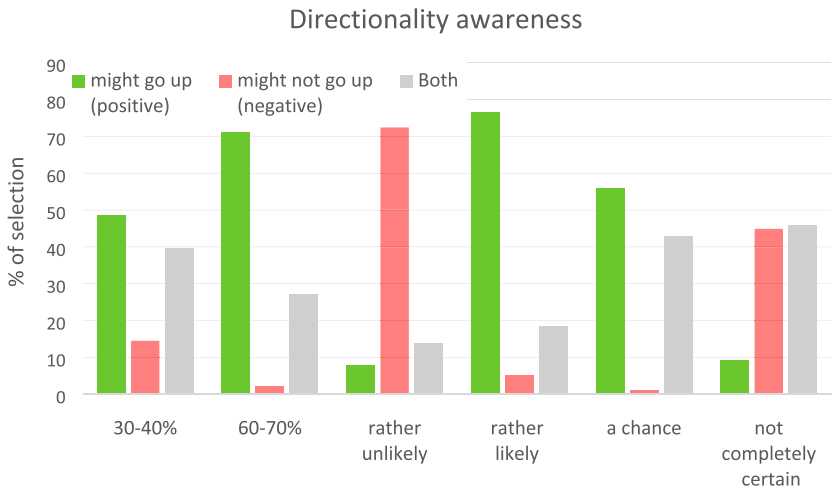

FIGURE 1 Interpretations of communicative meanings implied by numerical and verbal statements (percentages of participants in six conditions). When someone says, "it is … that the stocks will go up," do they try to direct your attention to the possibility that the stocks might go up or that they might not go up?

completely neutral but were perceived to direct a listener's attention primarily to the possibility that the stocks might go up, even in the case of low probabilities. Such "directionality" of numbers has previously been investigated by Bilgin and Brenner (2013) and by Teigen and Brun (2000). The two positive verbal phrases (*a chance* and *rather likely*) directed attention towards occurrences, whereas the two verbal negative phrases indicated that the target outcome might not occur. This is particularly conspicuous for the original unlikely/likely pair (middle panel), but even the two new ones (right panel) were perceived to direct attention upward and downwards, in line with their presumed directionality, rather than based on the probability they were supposed to convey. A focus on both alternative outcomes (may/may not happen) was most common for "30% to 40%," "a chance," and "not completely certain," which was particularly close to 50%.

4.2.2 | Effects of the second forecast

Will a second forecast produce a change in beliefs when it contains no new information, but just repeats the estimate of the first forecaster? As a measure of potential changes, we subtracted each participant's first rating (based on a single statement) from their second rating that was based on two identical statements (Judgment 2 − Judgment 1). A positive difference means that the second rating had strengthened their belief, a negative difference that it had weakened it. M&G found positive differences for "rather likely" and negative for "rather unlikely." They concluded that a combination of two verbal forecasts moved beliefs more closely towards the endpoints of the rating scale. For numerical probabilities, very few changes were observed in their study.

Table 1 shows how participants changed their beliefs after seeing two identical forecasts in each of six conditions. The first row shows that two equal numeric statements (including probabilities of 30% to 40%) made a stronger impression than one single estimate, in contrast to M&G who did not show a consistent reinforcement effect of the second numeric estimate. The magnitude of this change in the present study was not affected by probability level, $t(201) = 0.30$, *ns*. The second row in the table indicates an even stronger positive reinforcement effect of verbal phrases with positive directionality, whereas the bottom row displays a negative change for the two phrases with negative directionality. Again, the magnitude of change in the positive condition was not significantly affected by high versus low probability expression, $t(195) = 0.93$, $p = .18$, and the magnitude of change in the negative condition was similar for both negative phrases, regardless of the probability they indicated, $t(190) = 0.60$, $p = .28$. Data for likely and unlikely replicate the gist of M&G's study: Two "rather unlikely" statements made the forecasted event seem more unlikely (a negative difference) whereas two "rather likely" forecasts increased the probability of the same event (a positive difference), $t(196) = 7.99$, $p < .001$, $d = 1.14$.

Unlike M&G we predicted that the direction of a change was due to directionality, rather than level of probability. This was confirmed by the novel set of verbal phrases. "A chance" was believed to indicate a lower probability than "not completely certain" but is positively framed. It turned out that the ratings of these two VPEs were quite similar. (In fact, the ratings of the negative expression were slightly lower than the positive expression, but not significantly so.) As a result, when both advisors said there is "a chance" that the value of the LPT stocks would rise, it was more strongly believed (a positive difference) than when "a chance" was based upon a single forecast. In contrast, two "not completely certain" forecasts appeared *less* likely than one single forecast, $t(194) = 7.66$, $p < .001$, $d = 1.10$. An overall 2×2 ANOVA for the four verbal conditions yielded a main effect of original versus novel set of phrases, $F(1, 397) = 6.52$, $p = .011$, $\eta_p^2 = .016$, and of probability level $F(1, 397) = 34.49$, $p < .001$, $\eta_p^2 = .080$, and most importantly, a predicted interaction, $F(1, 397) = 39.20$, $p < .001$, $\eta_p^2 = .090$, indicating that the reinforcement effect was different for the novel set of phrases, where the lower probability phrase ("a chance") *strengthened* the likelihood when repeated. Both pairs of VPEs yielded more conspicuous directionality effects than numbers did.

A 2×2 ANOVA of format (verbal vs. numeric) and level of probability (high vs. low) for the original VPEs (rather likely and rather

TABLE 1 Mean ratings (1–10) of the probability of stock rise, based upon one vs. two equal forecasts in Study 1 (standard deviations in parentheses)

	Lower probability expressions			Higher probability expressions		
Condition	**Based on first forecast**	**Based on both forecasts**	**Change**	**Based on first forecast**	**Based on both forecasts**	**Change**
Numerical	30% to 40%		**0.55*****	60% to 70%		**0.50*****
	4.77 (1.40)	5.32 (1.95)		6.71 (0.89)	7.21 (1.19)	
Verbal positive	A chance		**0.86*****	Rather likely		**1.01*****
	5.91 (1.24)	6.77 (1.49)		7.03 (1.28)	8.04 (1.49)	
Verbal negative	Rather unlikely		**−0.66*****	Not completely certain		**−0.52*****
	4.15 (1.72)	3.49 (2.12)		5.49 (1.74)	4.97 (1.93)	

Note: Reinforcement effects in bold.
$^{***}p < .001$ (paired samples *t* tests); effect sizes (Cohen's *d*) between .37 and .93.

unlikely) yielded a significant interaction, $F(1, 397) = 39.20$, $p < .001$, $\eta_p^2 = .09$, which indicates that combinations of two high positive and two low verbal negative expressions are more different from each other than high and low numbers, replicating M&G's findings. A corresponding 2 × 2 ANOVA of format (verbal vs. numeric) and level of probability (high vs. low) for the novel pair of verbal phrases also revealed also a highly significant interaction, $F(1, 395) = 27.22$, $p < .001$, $\eta_p^2 = .064$, but in the opposite direction. Thus, positive verbal expressions that reflect probabilities closer to 50% can be "added" into something more likely, whereas negative expressions are combined into something less likely.

To examine more closely whether the changes depended on likelihood judgments instead of, or in addition to, directionality, we analyzed separately the judgments of participants who initially gave low probability ratings (1–5) or high ratings (6–10) to "a chance" and "not completely certain." It turned out that both groups produced a similar number of positive changes, as shown in Figure 2. For "a chance," the figure shows no instance of low ratings that led to *lower* ratings when combined, as would have been predicted by M&G's probability-based extremity hypothesis. Conversely, a combination of two "not completely certain" produced more downward than upward changes, both for ratings below (1–5) and above (6–10) the scale midpoint, as shown in the right part of Figure 2.

Thus, we partly replicated M&G's finding that two likely estimates make a prediction more likely and two unlikely statements make predictions less likely compared to only one, and in addition that this effect is stronger for verbal estimates than for numeric ones. But in contrast to these authors, we showed that the direction of the reinforcement effect is due to directionality and not to extremity of probability.

5 | STUDY 2: ONE VERSUS TWO UNEQUAL PROBABILITIES

This study was designed to extend the investigation to situations where predictions from two sources differ from each other. Independent stock market analysts draw conclusions about trends that often are at variance, one being perhaps more conservative and the other more optimistic. When their predictions are expressed in terms of numbers, we might expect observers to infer a compromise solution, placing the combined likelihood on a rating scale between the two separate estimates. For verbal probabilities, we hypothesized, based on results from Study 1, that two positive phrases (e.g., *a chance* and *rather likely*) would strengthen each other and add up to something higher than their average, whereas an exposure to two negative phrases (e.g., *rather unlikely* and *not completely certain*) would reinforce the possibility of a failure and make a judge more doubtful than warranted by the average of the two statements.

5.1 | Method

5.1.1 | Participants

Participants were recruited from Prolific with quotas on gender, age, and ethnicity to be representative of the UK population. After excluding 12 participants who failed the attention check and/or did not complete the questionnaires, the final sample consisted of 477 participants (246 women, 231 men, 0 other), with ages ranging from 18 to 81 years ($M = 44.6$, $SD = 15.2$). All but 2.1% had completed high school, and 59.5% reported to have obtained a bachelor's degree or higher.

5.1.2 | Design

All participants read two probabilistic forecasts, one high and one low, in counterbalanced order. The forecasts were either numerical, verbal positive, or verbal negative, in three between-subjects conditions.

5.1.3 | Questionnaires

All participants received the stock scenario from Study 1, but the two advisors did not make the same forecast. In the *Numerical* condition,

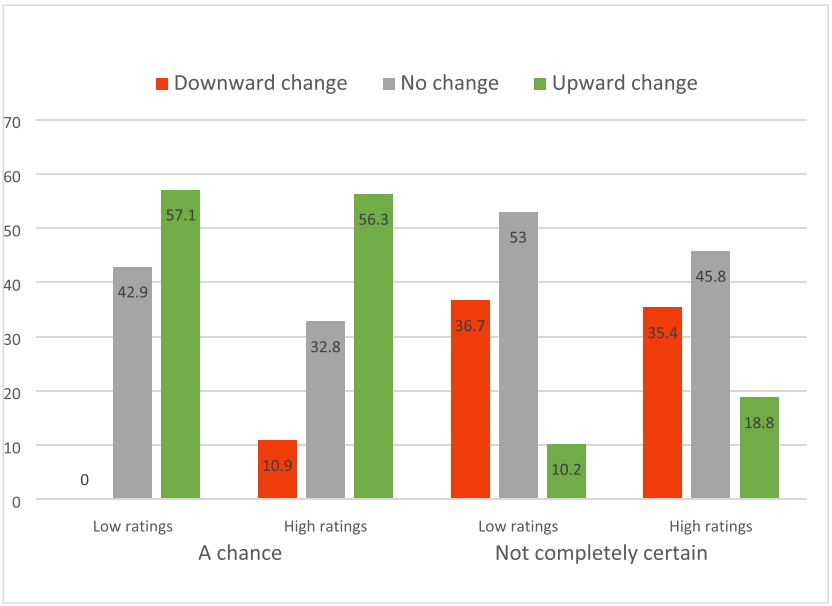

FIGURE 2 Percentages of participants who changed their beliefs upwards or downwards after receiving identical forecasters from a second advisor, split according to their initial ratings: low (1–5) or high (6–10) for "a chance" and "not completely certain", Study 1

one analyst said the LPT stocks had a 30% to 40% probability to increase, whereas the other gave a 60% to 70% estimate. In the *verbal positive* condition, Analyst A said: "a chance" and B: "rather likely," and in the *verbal negative* condition, A said: "rather unlikely" and B: "not completely certain" (or vice versa, resulting in an ascending and a descending order within each pair).

Participants rated their perceptions of the future profitability of these stocks on a 1–10 rating scale, as in Study 1, first based only on Analyst A's estimate, and then on the predictions of both analysts combined.

5.2 | Results

Mean ratings after one versus two unequal forecasts are presented in Table 2. A comparison of single forecasts (the two columns for "Based on first forecast") shows, as expected, that the presumed low forecast (e.g., 30% to 40%) were in all conditions rated clearly lower than the presumed higher forecast (e.g., 60% to 70%). After the second forecast, all ratings were adjusted towards the corresponding higher [lower] forecast in the pair. However, these adjustments were in all conditions so large that the final ratings, which were supposed to reflect both advisors' estimates, approached the value of the initial single ratings of these expressions. This strong order effect, which is evident from a comparison between the ascending versus descending order of ratings, was not expected, since both sets of ratings were presumably based on the same two statements, nothing being said about their temporal order, or whether the second statement in the pair should be regarded as more authoritative than the first one. The order effect was present in all conditions, but not to the same extent.

To test a potential difference between the adjustments in the positive and negative verbal conditions, we subtracted each participant's first rating (based on a single statement) from their second rating, which was based on both statements (Judgment 2 − Judgment 1), as in Study 1. A positive score here indicates a stronger belief after the second rating, while a negative score indicates a weaker belief, as shown in the columns for "Change" in Table 2. A 2 × 2 ANOVA for the verbal conditions, with directionality (positive vs. negative) and order (low–high vs. high–low) as between-subjects factors, showed a strong effect of order, $F(1, 314) = 206.58$, $p < .001$, $\eta_p^2 = .40$, and an effect of directionality $F(1, 314) = 4.01$, $p = .046$, $\eta_p^2 = .01$, with no interaction, $F(1, 314) = 1.80$, $p = .181$, $\eta_p^2 = .01$.

Comparing each verbal condition with the numerical condition (in separate 2 × 2 ANOVAs) showed an interaction between format and order, both for positive expressions, $F(1, 313) = 14.49$, $p < .001$, $\eta_p^2 = .04$, and for negative expressions, $F(1, 315) = 6.35$, $p = .012$, $\eta_p^2 = .02$; with numerical probabilities leading to stronger upward and downward changes than corresponding positive and negative verbal expressions. This could arguably be due to the greater numerical difference between high and low numerical probabilities (a difference of 30 percentage points) and is not in line with the idea that people use an averaging strategy for numerical probabilities.

5.3 | Discussion

Study 2 provided additional evidence for the reinforcement effect in the context of unequal verbal probabilities based on directionality rather than their degree of probability. But this effect was overshadowed and partly masked by a much stronger effect of presentation order, in that the statements of the second advisor apparently dictated the combined rating more strongly than the first advisor. We propose four different explanations of this puzzling phenomenon. (i) Participants misunderstood the task and rated only the prediction of the second advisor when they gave their second judgment. (ii) The second advisor was placed in the role of an arbiter, whose "second

TABLE 2 Mean ratings (1–10) of probability of stock rise, based upon one vs. two unequal forecasts in Study 2 (standard deviations in parentheses)

	Ascending order of ratings (lower expression first)			Descending order of ratings (higher expression first)		
Condition	Based on first forecast	Based on both forecasts	Change	Based on first forecast	Based on both forecasts	Change
Numerical (30% to 40% and 60% to 70%)	4.38 (1.58)	6.13 (1.50)	**1.75**	6.50 (1.05)	4.70 (1.10)	**−1.80**
Verbal positive (a chance and rather likely)	5.44 (1.53)	6.76 (1.43)	**1.32**	6.80 (1.61)	5.82 (1.49)	**- .98**
Verbal negative (rather unlikely and not completely certain)	3.46 (1.28)	4.66 (1.20)	**1.20**	5.30 (1.44)	3.74 (1.34)	**−1.56**

Note: Reinforcement effects in bold.

opinion" carried more weight than the original estimate (people ask for second opinions mainly when they are not happy with the first one). But the instructions clearly said that the rating should be based on both advisors' forecasts, and even if some participants misunderstood or elaborated the task, they would hardly be so numerous and so consistent in their responses as to produce a massive effect of order. (iii) It must have been quite obvious to most participants that they had to modify their first rating when presented with a second higher or lower estimate. To indicate a modified rating, one scale point was the smallest adjustment they could make. Accordingly, they ended up with an adjusted score after the second advisor which proved to be, on average, one scale point higher or one point lower than the first one, in line with the changes reported in Table 2. (iv) Finally, the change from the first advisor to the second might suggest a "trend," even if the presentation order did not imply a temporal sequence. Recent studies demonstrate that forecasts that are changed are readily perceived to indicate a *trend* continuing into the future (Erlandsson et al., 2018; Hohle & Teigen, 2015, 2019) perhaps reflecting a kind of psychological *momentum* (Maglio & Polman, 2016).

6 | STUDY 3: SEPARATE VERSUS COMBINED JUDGMENTS

In the previous studies, the same participants made a first judgment based on a single forecast, and then a second judgment based on two forecasts, leading to a revision of their original judgments. In contrast to this within-subjects procedure, the third study used a between-subjects design. Participants in a *separate* judgment condition estimated the probability equivalents of single forecasts, whereas other participants in a *combined* condition based their judgments on two forecasts. We hoped in this way to avoid the strong order effect that came when both tasks were performed by the same participants. The judgments were expressed as 0% to 100% numeric probabilities rather than ratings on a 1–10 rating scale. We hypothesized, as before, that two verbal expressions with a positive directionality (*a chance* and *rather likely*) would reinforce each other to yield a higher perceived probability than a simple average of the two individual terms, whereas two expressions with negative directionality would reinforce each other downward to yield a lower expectation of profitability.

6.1 | Method

6.1.1 | Participants

Participants were recruited from Prolific with quotas on gender, age, and ethnicity to be representative of the UK population. After excluding 17 participants who failed the attention check, the final sample consisted of 920 participants (500 women, 480 men, 12 other), with ages ranging from 18 to 87 years ($M = 45.4$, $SD = 15.5$). Altogether 55.9% reported to have obtained a bachelor's degree or equivalent. The stock vignette was appended to a questionnaire about an unrelated theme.

6.1.2 | Design

Participants were randomly assigned to one of six conditions in a 2×3 design, where one factor was separate versus combined judgments, and the other numerical versus verbal positive versus verbal negative expressions of probability, as in the other studies.

6.1.3 | Questionnaire

All participants received a vignette similar to the one used in Study 2, describing two analysts' forecasts about stocks, with only minor changes in wording to enhance readability (e.g., naming analysts and asking them about "profitability" rather than stocks that close above a certain value).

In the three *single judgment* conditions, two analysts, Paula and Tracey, made forecasts about two different companies, Alpha and Epsilon. The two forecasts were shown on the same page, with one forecast conveying a low probability and the other a high probability, with the order being counterbalanced across participants, resulting in an ascending and a descending order within each pair. In the *verbal positive* condition, Paula said that Alpha stocks had "a chance" of being profitable, whereas Tracey said: "rather likely" for Epsilon; in the *verbal negative* conditions Paula said: "rather unlikely" and Tracey said: "not completely certain" (or vice versa). In the *numerical* condition, one analyst said that Alpha stocks had a 30% to 40% probability

to be profitable, whereas the other gave a 60% to 70% estimate for Epsilon. For each forecast, participants were asked to give their own probability estimate by moving a cursor on a slider from 0% to 100%. This scale was used instead of the 1–10 rating scale of Studies 1 and 2 to facilitate a probabilistic mindset.

In the three *combined judgments* conditions, participants were informed about two analysts giving different forecasts about the same stocks (Alpha). They used the same pairs of positive verbal probabilities, verbal negative probabilities or numerical probabilities as in the single judgment tasks above, and the order of the terms was also counterbalanced. Their task was only to give their own probability estimate of Alpha's profitability based on both analysts' estimates in combination.

Participants in the combined conditions were finally asked to rate the perceived agreement of the analysts on a 5-point Likert scale from 1: *strongly disagree* to 5: *strongly agree*. We expected that two VPE with the same directionality would be perceived as indicating agreement even when their probabilities are different, whereas agreement would be less for probabilities conveyed in a numeric format. We also hypothesized that a strengthened forecast—a weaker statement (e.g., "a chance") followed by a stronger statement (e.g., "rather likely")—might be perceived as more in agreement than the same two forecasts presented in the opposite order (the second "weakening" the first one).

6.2 | Results

6.2.1 | Combined versus single probabilities

Participants in the single forecast conditions gave estimates that roughly replicated the probability estimates from Study 1. All low probability expressions (a chance, rather unlikely, and 30% to 40%) were believed to indicate probabilities below 50%, and high probability expressions indicated probabilities from 50% and upwards, as shown in the first two columns of Table 3. The third column displays the means of both expressions in each pair, which would be the expected compromise estimate in cases where a pair of analysts suggested different probability estimates. But results from the combined forecast conditions (fourth column in the table) reveal that people do not simply average the analysts' single estimates but suggest estimates that are skewed towards a higher probability in the verbal positive condition and towards a lower probability in the verbal negative condition. We find accordingly evidence of a positive reinforcement effect for verbal probabilities that have a positive directionality and a negative reinforcement effect for verbal probabilities that have a negative directionality, even for unequal VPEs. Two numeric forecasts (first row in Table 3) were more often averaged, but even in this case, there was a slight, consistent tendency to estimate the combined forecast above the mean of the two single forecasts.

6.2.2 | Agreement judgments

Do analysts who say that "there is a chance" or "it is rather likely" agree or disagree about the future profitability of Alpha stocks? A previous study by Løhre et al. (2019) showed that perceived agreement between forecasters is affected by the way their estimates are framed, for instance in terms of occurrence or non-occurrence of a specific outcome. Directionality of verbal terms can be regarded as an instance of framing (Teigen & Brun, 2003). The mean agreement ratings displayed in Figure 3 indicate that those using verbal expressions are perceived to agree more than they disagree (scores above the "neutral" scale midpoint of 3.0), whereas those using high versus low numeric probabilities are perceived to be in disagreement with each other (mean scores below 3.0). A 3×2 ANOVA with condition and order as the two factors revealed a highly significant effect of condition (numeric vs. verbal positive vs. verbal negative), $F(2, 491) = 27.32$, $p < .001$, mainly due to the large difference between the numeric and the two verbal conditions. There is also an effect of order (low p first vs. high p first), $F(1, 491) = 4.17$, $p = .042$, supporting that two estimates appear to be slightly more in agreement with each other when presented in ascending than in descending order.

7 | STUDY 4: SELECTING VERBAL PROBABILITIES

In Study 1, we found reinforcement effects for likelihood judgments performed on a neutral (unlabeled) rating scale. But M&G had added labels (*rather likely*, *quite likely*, etc.) to each scale value, suggesting

TABLE 3 Mean judged probabilities of stock rise, based upon two unequal forecasts in Study 3 (0% to 100% scale, standard deviations in parentheses)

	Single forecasts					Test of diff	
Condition	Lower probability	Higher probability	Mean of two	Combined forecast	Diff	*t*	*p*
Numerical (30% to 40% and 60% to 70%)	34.8 (9.5)	60.8 (9.0)	47.8 (8.5)	49.9 (7.4)	**2.1**	2.43	<.01
Verbal positive (a chance and rather likely)	41.1 (16.7)	59.9 (13.9)	50.3 (11.9)	54.3 (15.6)	**4.0**	2.60	<.01
Verbal negative (rather unlikely and not quite certain)	29.8 (15.3)	50.0 (16.5)	39.6 (11.9)	33.2 (14.3)	**−6.4**	4.42	<.001

Note: The tests of difference compare the combined forecasts with the mean of two (Column 4 − Column 3). Reinforcement effects in bold.

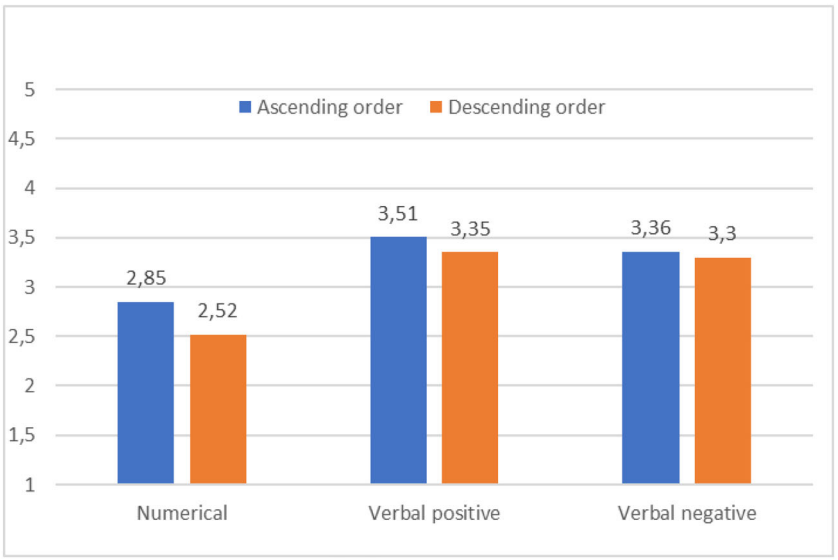

FIGURE 3 Mean perceived agreement between two analysts assessing probabilities with numerical and verbal expressions on Likert scales from 1: *strongly disagree* to 5: *strongly agree*. Data based on agreement scores from participants in the combined conditions, Study 3

that these verbal labels described the changes more directly. To capture whether such changes depended on directionality, an additional study was performed where participants were given a choice between different verbal expressions (labels) instead of performing numerical ratings. We used a pair of verbal probabilities (*a chance* vs. *not certain*), which supposedly indicate a similar probability with opposite directionality. When two advisors both used a positive phrase, "there is a chance," for Alpha stocks of being profitable, would you conclude that the combined chance is higher (e.g., "a fair chance")? Or if they both used a negative phrase, "not certain," does this suggest that the combined probability is lower (e.g., "quite uncertain")? We used in this experiment a coarse response scale, with only three levels (in contrast to the more fine-grained rating scales of the previous studies), to test the effect in a minimal design, where people could choose the same phrase as the advisors, or one conveying a higher or a lower probability.

7.1 | Method

7.1.1 | Participants

Participants were recruited from Prolific, as part of a brief questionnaire on "Predictions about quantities." The questions were answered by 200 participants (99 women, 99 men, and 2 other), with ages ranging from 18 to 81 years ($M = 38.6$, $SD = 14.8$). Almost all (94.5%) were native English speakers, and 47.0% reported to have a bachelor's degree or equivalent.

7.1.2 | Design

Participants were randomly assigned to two experimental groups, one (A) with positive verbal statements and the other (B) with negative ones.

7.1.3 | Questionnaires

Both groups read a scenario about two financial analysts who independently of each other predicted the future performance of the company Alpha. In Group A, both experts said: "There is *a chance* that Alpha stocks will be profitable." In Group B, both said: "It *is not certain* that Alpha stocks will be profitable." The VPEs were chosen to represent one positive and one negative phrase that capture a range of intermediate probability values.

All participants were asked: Based on these two analysts' predictions, what do you think is the probability that Alpha stocks will be profitable? Group A participants received a choice between three positive verbal phrases: *a fair chance*, *a chance*, or *a small chance*. Group B participants could choose between three negative phrases: *not completely certain*, *not certain*, or *quite uncertain*. It was assumed, and subsequently confirmed by a ranking procedure, that the first phrase in each set conveyed a higher probability, and the third phrase described a lower probability than the phrase in the middle.[1] They were finally asked whether they considered the two verbal terms (*a chance* or *not certain*) as conveying probabilities below 50%, about 50%, or above 50%.

7.2 | Results

When both financial analysts said there was "a chance" that Alpha stocks were profitable, most participants also chose "a chance" as their preferred judgment. When both analysts said it was "not certain," most participants adopted "not certain" as their preferred response, as shown in Table 4. This was expected since both "a chance" and "not certain" are quite broad and could be used about a range of uncertain events and could accommodate (and mask) most minor changes. Besides, just copying the advisors' phrases would be the most obvious and effortless "default" response. However, among those that did not simply repeat the "neutral" middle option, we

TABLE 4 Number of participants (percentages) who combined two equal verbal statements by selecting a higher, an intermediate (unchanged) or a lower VPE, Study 4

Condition	Higher	Intermediate	Lower	Total
A: Verbal positive	A fair chance:	A chance:	A small chance:	
	27 (27.5%)	63 (64.3%)	8 (8.2%)	98 (100%)
B: Verbal negative	Not completely certain:	Not certain:	Quite uncertain:	
	3 (2.9%)	90 (88.2%)	9 (8.8%)	102 (100%)

expected a directionality effect. Table 4 shows that the patterns of answers in the positive and negative conditions diverged from each other, $\chi^2(2) = 23.95$, $p < .001$. Participants in the verbal positive condition selected a higher probability expression ("a fair chance") more often than a lower ("a small chance"), whereas this pattern was reversed for negative terms. A crosstabs analysis of upward versus downward changes yielded a $\chi^2(1) = 10.52$, $p = .001$, Cramer's $V = .47$. Controlling for probability, we also performed this analysis on the subset of participants who estimated "a chance" and "not certain" to mean probabilities of about 50%. This analysis yielded $p = .011$ by Fisher exact test, indicating that the reinforcement effect was not primarily due to probability level but to the directionality of the verbal phrase.

8 | GENERAL DISCUSSION

We investigated how people combine two verbal probabilities depending on their probability magnitude and directionality. Recent work by Mislavsky and Gaertig (2021) demonstrated that verbal probabilities from two sources were combined in an additive way rather than averaged. Two identical low verbal probabilities led to a lower combined probability whereas two identical high ones led to a higher combined estimate. This reinforcement effect did not occur for numeric probabilities in their study. We aimed to replicate their observations and to identify the process at work, while extending the approach to pairs of probabilities that were not identical.

In Study 1, we replicated M&G's finding that two identical verbal probability forecasts can reinforce each other, both in an upward and in a downward direction. We found the same effect as M&G for "likely" and "unlikely," but unlike these investigators, we attributed the effect to the directionality of the verbal expressions, rather than to the magnitude of the probabilities conveyed. A combination of two low-level *positive* probability forecasts ("a chance") was perceived as *more* likely, whereas two high-level *negative* forecasts ("not completely certain") led to a *less* likely judgment than a single statement.

Our study showed in addition that numeric estimates were not immune to an upward reinforcement effect, as several participants rated the combination of two numeric statements more likely than either of them (Tables 1 and 3). This may be due to a minor but important procedural difference. M&G had provided precise numeric labels for the units on the rating scale, which made it very easy to tick the box for 30% to 39% when both advisors had made this exact estimate. In our studies, we used more approximate ranges (30% to 40%) and did not repeat these labels on the response scales, allowing participants more latitude in their responses.

The directionality of verbal phrases does not just allow us to predict what happens when verbal forecasts are combined, it brings us also closer to an explanation of this effect. Previous research has shown that positive and negative phrases are not neutral but perceived as relative to a reference point (Honda & Yamagishi, 2017; Juanchich et al., 2010). Like other instances of framing, directionality triggers specific inferences: A 'half-full' glass will be regarded as previously empty, whereas a "half-empty" glass was previously full but is now partially emptied (McKenzie & Nelson, 2003). Similarly, a "rather likely" event will be contrasted with an unlikely outcome, whereas a "not completely certain" outcome is contrasted with a certain one. This "dynamic" feature of verbal probabilities, pointing either upwards or downward, might explain why verbal probabilities are perceived as expressing opinions and giving advice (Teigen & Brun, 1999), less ambiguously so than corresponding numerical probabilities (Collins & Mandel, 2019). Stocks that have "a chance" to rise may not perform so well right now but may have a future potential. It follows that two advisors using positive (equal or unequal) verbal phrases are perceived to "join forces" in a more compelling way than if they just were perceived as handing out the same neutral piece of information.

Studies 2 and 3 provided additional evidence for the reinforcement effect in a context of unequal verbal probabilities (e.g., a low and a high probability instead of two identical expressions). In Study 2, the effect was weaker, being overshadowed by an effect of presentation order, in that the statements of the second advisor apparently dictated the combined rating much more strongly than the first advisor. To minimize the order effect, Study 3 was conducted with a between-subjects design, where different participants performed single or combined probability judgments. This study demonstrated a reinforcement effect of directionality also for unequal verbal probabilities. When one advisor says: "there is a chance" and another estimates this outcome to be "rather likely," the higher probability will prevail. For negative probabilities, the opposite occurs. A "rather unlikely" and "not completely certain" forecast are not simply averaged but produce a combined probability that is biased towards the lower of these two. Further studies should be conducted to explore how positive and negative expressions are combined. We hypothesize that such statements are perceived to be in disagreement and hence less likely to reinforce each other.

Overall, across studies, we find the largest format differences between numeric and negative verbal phrases, whereas numbers and positive verbal phrases are more similar. Thus, the format difference in combining forecasts is in the present studies less prominent than Mislavsky and Gaertig's (2021) data suggested. Although "60% and 60% is 60%," the combination may be rated as something more than 60%, and "likely and likely" is not always "very likely" (as the title of the original study suggested), but just a bit more likely combined than when assessed separately. The similarity of verbal positive and numeric probabilities found in the present studies should not be taken as argument for words to be an imperfect kind of numbers in disguise, but rather as evidence for the "directionality of numbers" (Bilgin & Brenner, 2013; Teigen & Brun, 2000). A "60% probability" is not just an abstract number, it has a verbal referent, namely, *probability*, which makes it directed more towards occurrences than non-occurrences. Participants in Study 1 agreed that a speaker who said the stocks had a 60% to 70% probability and even when they had 30% to 40% probability to rise, directed their attention towards the possibility that they might go up.

The studies extend our understanding of what is conveyed by verbal expressions of probability. The upward or downward directionality of such expressions have earlier been shown to direct recipients' attention towards different potential outcomes, implying different reference points (Honda & Yamagishi, 2017) and revealing the speaker's concerns and communicative intentions (Collins & Hahn, 2018; Collins & Mandel, 2019; Teigen & Brun, 1995, 1999). The present research shows that directionality also affects the inferences recipients draw when combining forecasts from different sources.

8.1 | Conclusion

We investigated the role of directionality in combining verbal probabilities, for two forecasts of the same event. Prescriptive models recommend a conservative, averaging approach (Armstrong et al., 2015), unless the forecasts are based on information from different and independent sources (Baron et al., 2014). Mislavsky and Gaertig (2021) suggested that the interpretations of such combinations are a product of the magnitude of the probabilities involved, whereas we find that the way verbal probabilities are framed, their directionality, is the decisive factor. Although all statements about uncertain events are Janus faced, suggesting both occurrence and non-occurrence of an outcome, some terms—the positive ones—are mutually strengthening each other, suggestive of a psychological momentum (Maglio & Polman, 2016), whereas others—those containing negations—will make an outcome appear increasingly doubtful.

DATA AVAILABILITY STATEMENT

Data files for all studies are available upon request.

ORCID

Karl Halvor Teigen https://orcid.org/0000-0001-6165-9034
Marie Juanchich https://orcid.org/0000-0003-0241-9529

ENDNOTE

[1] A study of conjunctions (not reported here) gave the following mean ranks for four positive statements from most certain to least certain: A fair chance = 1.51; a chance = 1.80; a small chance = 2.91; a very small chance = 3.79. Mean ranks for four negative statements from most certain to least certain: not completely certain = 1.67; not certain = 2.24; quite uncertain = 2.53; very uncertain = 3.57.

REFERENCES

Armstrong, J. S., Green, K. C., & Graefe, A. (2015). Golden rule of forecasting: Be conservative. *Journal of Business Research*, *68*, 1717–1731. https://doi.org/10.1016/j.jbusres.2015.03.031

Baron, J., Mellers, B. A., Tetlock, P. E., Stone, E., & Ungar, L. H. (2014). Two reasons to make aggregated probability forecasts more extreme. *Decision Analysis*, *11*(2), 133–145. https://doi.org/10.1287/deca.2014.0293

Bilgin, B., & Brenner, L. (2013). Context affects the interpretation of low but not high numeric probabilities: A hypothesis testing account of subjective probability. *Organizational Behavior and Human Decision Processes*, *121*, 118–128. https://doi.org/10.1016/j.obhdp.2013.01.004

Collins, P. J., & Hahn, U. (2018). Communicating and reasoning with verbal probability expressions. *Psychology of Learning and Motivation*, *69*, 67–105. https://doi.org/10.1016/bs.plm.2018.10.003

Collins, R. N., & Mandel, D. R. (2019). Cultivating credibility with probability words and numbers. *Judgment and Decision Making*, *14*(6), 683–695.

Erlandsson, A., Hohle, S. M., Løhre, E., & Västfjäll, D. (2018). The rise and fall of scary numbers: The effect of perceived trends on future estimates, severity ratings, and help allocations in a cancer context. *Journal of Applied Social Psychology*, *48*(11), 618–633. https://doi.org/10.1111/jasp.12552

European Food Safety Authority, Hart, A., Maxim, L., Siegrist, M., von Goetz, N., da Cruz, C., Merten, C., Mosbach-Schulz, O., Lahaniatis, M., Smith, A., & Hardy, A. (2018). Guidance on communication of uncertainty in scientific assessments. *EFSA Journal*, *17*(1), e05520. https://doi.org/10.2903/j.efsa.2019.5520

Fisk, J. E. (2022). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment, and memory* (3rd ed., pp. 27–43). Psychology Press. https://doi.org/10.4324/9781003154730-4

Gurmankin, A. D., Baron, J., & Armstrong, K. (2004). The effect of numerical statements of risk on trust and comfort with hypothetical physician risk communication. *Medical Decision Making*, *24*(3), 265–271. https://doi.org/10.1177/0272989X04265482

Hohle, S. M., & Teigen, K. H. (2015). Forecasting forecasts: The trend effect. *Judgment and Decision Making*, *10*(5), 416–428.

Hohle, S. M., & Teigen, K. H. (2019). When probabilities change: Perceptions and implications of trends in uncertain climate forecasts. *Journal of Risk Research*, *22*, 555–569. https://doi.org/10.1080/13669877.2018.1459801

Honda, H., & Yamagishi, K. (2017). Communicative functions of directional verbal probabilities: Speaker's choice, listener's inference, and reference points. *The Quarterly Journal of Experimental Psychology*, *70*(10), 2141–2158. https://doi.org/10.1080/17470218.2016.1225779

Juanchich, M., Sirota, M., & Bonnefon, J.-F. (2019). Verbal uncertainty. In C. Cummins & N. Katsos (Eds.), *Oxford handbook of experimental semantics and pragmatics*. Oxford Press. https://doi.org/10.1093/oxfordhb/9780198791768.013.2

Juanchich, M., Teigen, K. H., & Villejoubert, G. (2010). Is guilt "likely" or "not certain"? Contrast with previous probabilities determines choice of verbal terms. *Acta Psychologica*, *135*(3), 267–s. https://doi.org/10.1016/j.actpsy.2010.04.016

Løhre, E., Sobków, A., Hohle, S. M., & Teigen, K. H. (2019). Framing experts' (dis)agreements about uncertain environmental events. *Journal of Behavioral Decision Making*, *32*, 564–578. https://doi.org/10.1002/bdm.2132

Maglio, S. J., & Polman, E. (2016). Revising probability estimates: Why increasing likelihood means increasing impact. *Journal of Personality and Social Psychology*, *111*(2), 141–158. https://doi.org/10.1037/pspa0000058

Mandel, D. R., & Irwin, D. (2021). Uncertainty, intelligence, and national security decision-making. *International Journal of Intelligence and CounterIntelligence*, *31*, 558–582. https://doi.org/10.1080/08850607.2020.1809056

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., & Zwiers, F. W. (2010). *Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties*. Intergovernmental Panel on Climate Change (IPCC). http://www.ipcc.ch

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals. Reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review*, *10*, 596–602. https://doi.org/10.3758/BF03196520

Mislavsky, R., & Gaertig, C. (2021). Combining probability forecasts: 60% and 60% is 60%, but likely and likely is very likely. *Management Science*, *68*(1), 541–563. https://doi.org/10.1287/mnsc.2020.3902

O'Brien, B. J. (1989). Words or numbers? The evaluation of probability expressions in general practice. *Journal of the Royal College of General Practitioners*, *39*, 98–100.

Piercey, M. D. (2009). Motivated reasoning and verbal vs. numerical probability assessment: Evidence from an accounting context. *Organizational Behavior and Human Decision Processes*, *108*(2), 330–341. https://doi.org/10.1016/j.obhdp.2008.05.004

Smith, S. W., & Jucker, A. H. (2014). "Maybe, but probably not": Negotiating likelihood and perspective. *Language and Dialogue*, *4*, 284–298. https://doi.org/10.1075/ld.4.2.06smi

Teigen, K. H. (1988). The language of uncertainty. *Acta Psychologica*, *68*, 27–38.

Teigen, K. H., & Brun, W. (1995). Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, *88*, 253–258.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes*, *80*, 155–190. https://doi.org/10.1006/obhd.1999.2857

Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: When does p = 0.3 reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making*, *13*, 345–362. https://doi.org/10.1002/1099-0771(200007/09)13:3<345::AID-BDM358>3.0.CO;2-U

Teigen, K. H., & Brun, W. (2003). Verbal probabilities: A question of frame? *Journal of Behavioral Decision Making*, *16*, 53–72. https://doi.org/10.1002/bdm.432

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

How to cite this article: Teigen, K. H., Juanchich, M., & Løhre, E. (2022). Combining verbal forecasts: The role of directionality and the reinforcement effect. *Journal of Behavioral Decision Making*, 1–13. https://doi.org/10.1002/bdm.2298