Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Cry Wolf or Equivocate? Credible Forecast Guidance in a Cost-Loss Game

Gary E. Bolton, Elena Katok

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 64, No. 3, March 2018, pp. 1440–1457
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Cry Wolf or Equivocate? Credible Forecast Guidance in a Cost-Loss Game

Gary E. Bolton,[a] Elena Katok[a]

[a] Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080
Contact: gbolton@utdallas.edu (GEB); ekatok@utdallas.edu (EK)

Received: October 14, 2014
Revised: December 18, 2015; April 14, 2016
Accepted: June 20, 2016
Published Online in Articles in Advance: January 5, 2017

https://doi.org/10.1287/mnsc.2016.2645

Abstract. What is the most credible way to convey the risk in expert forecasts to the nonexpert decision makers who use the forecast? We test two ways to communicate this information: provide an unequivocal recommendation or equivocate by providing the probability of the uncertain event of interest. We use a simple game in which human subjects (the forecast users) decide whether to take a risk of a loss or pay a cost to avoid the risk. We find that the influence of forecast information differs depending on whether the information implies that the decision maker should take the status quo action, the preforecast optimal action, or the siren action, optimal only if the period-by-period forecast implies it. Unequivocal recommendations are more successful at inducing the status quo action. Equivocating by providing probabilities is more successful at inducing the siren action. Unequivocal recommendations to take the siren action are less effective because the false certainty makes them more vulnerable to the "cry wolf" effect. A forecast that combines probabilities with recommendations captures most of the gains of the separate approaches. A follow-up study examines whether an explanation of how risk quantities inform recommendations can substitute for providing the risk quantities in the forecast. The study finds little evidence in support of this proposition.

History: Accepted by Yuval Rottenstreich, judgment and decision making.
Supplemental Material: Data are available at https://doi.org/10.1287/mnsc.2016.2645.

Keywords: cost-loss game • experimental economics • behavioral operations • risk and decisions

1. Introduction

Many business and personal decisions involving risk have a cost-loss structure: the critical question is whether or not to take a cost to avoid the risk. Making an informed decision requires an assessment of how much risk there is (Weber 1994, Fox and Tversky 1998).[1] For many practical problems, experts use formal models to forecast these risks while a separate set of nonexpert users make the decisions.

Forecast guidance derived from expert models is an increasingly common feature of modern life; widely socially available, it guides decisions big and small, from improving emergency responsiveness (Green and Kolesar 2004, Craft et al. 2005, Pinker 2007), to managing supply base diversification, to purchasing auto insurance. The forecast for all these specific applications is inherently uncertain. And, in fact, given the ready availability of "big data," the models created by business analysts, meteorologists, climatologists, epidemiologists, and economic and finance professionals, among others, routinely churn out quantitative estimates, such as probabilities, for the various events of interest.

The larger research question we address is how to design forecast guidance for nonexpert users. In theory, the critical information a decision maker needs is the probabilities over the relevant contingencies. In practice, the estimates of risk that currently reach end users are sometimes vague or nonexistent. Forecasters sometimes opt for providing only a best-guess point estimate (a single contingency). But by itself, a best-guess point estimate deprives users of the risk information critical to making the optimal decision. Concerns about the point estimate approach have driven a push to include more measures of risk in publicly available forecasts. For example, after some study, probabilistic information has become more readily available in weather forecasting (U.S. National Research Council 2006). Reporting standards are, however, uneven across fields, as emphasized by Silver (2012), who points to specific deficiencies in economic and political forecasting.

Providing the relevant probabilities is a straightforward way of conveying forecast uncertainty to end users. Previous studies find that providing quantitative measures of risk can improve the quality of decisions relative to only providing best-guess point forecasts (Papastavrou and Lehto 1996), although quality falls short of the maximum benefits possible (Roulston et al. 2006). Decision makers have limited ability to process information (Kahneman 2003), including risk information represented numerically (Erev et al. 1993), so there are reasons to believe that this gap between theory and actual performance will persist.

In this paper, we report on an experiment that investigates the (relative) efficacy of providing risk information as a probability versus providing an unequivocal recommendation of the best course of action (based on the probability forecast). The recommendation approach ("You are in the path of a hurricane; you are advised to evacuate now.") is for the expert to both assess the risk information and draw conclusions about the appropriate course of action on behalf of the end user. Receiving forecast guidance in the form an unequivocal recommendation from an expert (e.g., financial analyst, economist, meteorologist, doctor or dentist) is a commonplace. It has the advantage of finessing the cognitive limitations surrounding the processing of quantitative information.

In the experiment we report here, each subject plays a series of cost-loss games. In each game, the decision maker, who is the forecast user, must choose between taking a risk of a *loss* with a certain probability P or avoiding the risk by paying a *cost*. Across treatments, we vary the magnitude of the cost, as well as the forecast information that we give our subjects. The forecast information takes the form of either the probability of the loss or an explicit recommendation of what action to take. The experiment includes treatments in which subjects have no (round-by-round) forecast information but only the long-term frequency of events. We study the influence of each kind of forecast information relative to having neither, as well as relative to one another. In addition, we study forecasts that combine probabilities with recommendations. Budescu et al. (2012) show that presenting probability ranges together with verbal probability terms improves user consistency in interpreting forecast guidance, so it is plausible that combining probabilities with recommendations might have a similar effect on the decisions studied here.

An important issue is how the credibility or trust that users put in the forecast changes as users gain experience with the forecast. There is an extensive literature to show that false alarms lead to less compliance with future alarms, a behavioral regularity known as the *cry wolf effect* (e.g., Breznitz 1984, Bliss et al. 1995, Meyer and Bitan 2002). A forecast that is ex post wrong or misleading can be thought of as false alarm (Roulston and Smith 2004). Studies of peoples' preferences for forecast information suggest how false alarms might lead to a cry wolf effect: preferences for forecast information exhibit an accuracy–informativeness trade-off, and on balance, decision makers tend to prefer forecasters that disclose the risk or uncertainty (Yaniv and Foster 1995, 1997). In fact, people prefer forecasts in which the communicated uncertainty matches the true uncertainty in the underlying decision context (Wallsten et al. 1993; see also Budescu and Wallsten 1987).

Because it does not communicate uncertainty, a recommendation that proves to be a false alarm might lead to a loss of trust in the forecaster and hence a cry wolf effect. Yet this reasoning might be on a continuum and apply to probability forecasts as well: in the game we study, each forecast probability implies a normative optimal action. A probability forecast can be taken as a false alarm when the implied action fails to be ex post correct. Even though a probability forecast, by its nature, communicates uncertainty, a false alarm might be viewed as evidence that the reported uncertainty fails to match the true uncertainty, leading to a loss of trust and a cry wolf effect. We test both recommendation and probability forecasts for cry wolf effects.

We also study how the framing of a forecast recommendation influences adherence to the forecast. Arkes et al. (1986) conduct a study in which subjects performed a task where the outcome had a random component. They find that subjects were more likely to adhere to a recommended rule after making incorrect judgments when warned that abandoning the rule was likely to lead to poorer performance (telling them that following the rule would result in a performance that was about as well as people can do had less effect). In our study, subjects are told that the recommendation is optimal in the sense that it maximizes the average expected payoff, and we study whether giving a more detailed explanation of the method behind this claim can improve adherence to the rule.

The experiment we report focuses on forecast probabilities in quantitative form. An extensive literature examines forecasts using verbal probability expressions. Verbal descriptions have the advantage of not suggesting overly precise estimations of uncertainty. At the same time, user interpretation of these terms is sensitive to context (e.g., Weber and Hilton 1990, Harris and Corner 2011) and often varies widely by the individual (e.g., Wallsten et al. 1986, Karelitz and Budescu 2004).

The main contribution of our work is a laboratory demonstration of (what turned out to be) a rather subtle interplay between probabilistic forecasting in cost-loss games and the format in which forecasts are presented to users (probabilities versus recommendations). To our knowledge, this relationship has never been demonstrated before in either the cost-loss game or cry wolf literatures. We find that recommendation forecasts are more effective at inducing the status quo action when it is optimal. A second finding is that probabilistic forecasts are more effective at inducing the siren action when it is optimal. Our third finding is that the cry wolf effect is more strongly associated with recommendation forecasts, which might explain why they are not as effective at inducing a siren action (note that the status quo action is optimal if one chooses to disregard the forecast).

In the next section we present the details of our experimental design, laboratory protocol, and analytical benchmarks. In Section 3 we present the results of the first study, starting with aggregate descriptive statistics, then individual-level analysis, and following with the dynamic analysis to determine the effect of the cry wolf effect. In Section 4 we present the results of the second study, again starting with descriptive statistics and following with the dynamic analysis. In Section 5 we summarize our conclusions and discuss managerial implications.

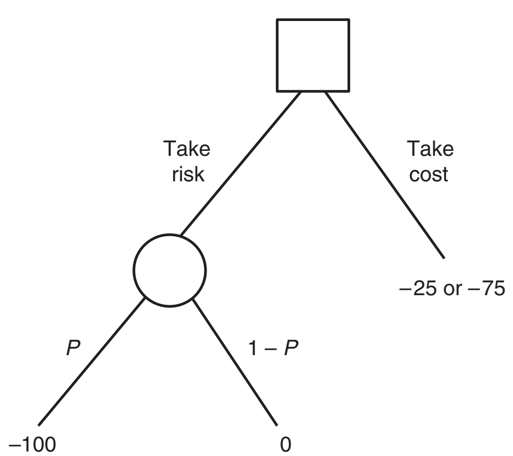
2. Design of Experiment and Materials

We display the extensive form of the cost-loss game graphically in Figure 1. The decision maker chooses between taking a risk of a loss with probability P or avoiding the risk by taking a cost. In all our treatments, *Loss* = 100 tokens. Some subjects play the game with a high cost (*Cost* = 75 tokens); others play with a low cost (*Cost* = 25 tokens).

The main condition manipulation is the information decision makers have about probability P, the value of which varies each game. At the beginning of each condition of the experiment, subjects are told that the probability of the loss event will average to 50% over the series of cost-loss games. In the baseline *Neither conditions*, this is all the information subjects are given. In the other conditions, subjects receive round-by-round forecast guidance. In the *Probability conditions*, forecast guidance is communicated as the actual value of P for that round. In the *Recommendation conditions*, guidance takes the form of advice: "take the risk" or "take the cost" (the value of P is not provided). Finally, in the *Both conditions*, both probability and recommendation are presented together.

In determining the expected-cost minimizing decision, one should weigh the cost of taking preventive action against the expected loss from no action.

Figure 1. The Cost-Loss Game



The expected value optimal decision for the cost-loss game is to take the risk as long as $P \times Loss < Cost$ or $P < Cost/Loss$. Otherwise, the expected value optimal decision is to take the cost. We will take the expected value optimal decision as our benchmark and hereafter refer to it as simply the *optimal* action.[2]

When the only available information is that the probability P averages 0.5, the optimal action for subjects facing the low cost treatment (*Cost* = 25) is to take the cost, while the optimal action for subjects facing the high cost treatment (*Cost* = 75) is to take the risk. When a round-by-round probability forecast is available, the optimal rule for subjects facing a low cost condition is to take the risk when $P < 0.25$ and take the cost otherwise. Subjects facing the high cost condition should take the risk when $P < 0.75$ and take the cost otherwise. The recommendation follows the same optimal decision rule, so that from the point of view of rational decision theory both probability and recommendation guidance should work equally well, whether provided separately or in combination.[3] The round-by-round forecast of the probability provides far better information about the risk than does knowing only the average probability, so decision makers with round-by-round forecasts should experience fewer total losses (costs paid plus losses incurred from risks) than users who only know the average probability.

We will refer to the action that is optimal when no round-by-round forecast guidance is available as the *status quo action*. In this context, the function of the round-by-round forecast is to alert the decision maker that the status quo action should be abandoned in favor of what we will call the *siren action*. For convenience, these definitions are laid out in Table 1.

Likewise, we can speak of a *siren forecast* and a *status quo forecast* depending on which action the forecast implies is optimal.

In our experiments, each decision maker completed 100 rounds of the cost-loss game. The forecast information given to the subject (Neither, Probability, Recommendation, or Both) depended on the treatment to which the subject had been randomly assigned, and this assignment remained constant throughout all rounds for each subject. Each subject participated in a single treatment.

We conducted two studies. In the first study we investigate the effect of recommendation and probability in two cost conditions (*Cost* = 25 and *Cost* = 75). In the second study we add an explanation to the directions

Table 1. Definition of Status Quo and Siren Actions, by Cost Condition

	Status quo	Siren
Low cost (25)	Take the cost	Take the risk
High cost (75)	Take the risk	Take the cost

given to subjects, describing how the recommendation is determined, and investigate whether this manipulation changes the interaction between recommendation and probability information, in the high cost condition ($Cost = 75$). At the beginning of each round, a subject was credited with 150 tokens. Individual round profit was then equal to 150 minus the cost or the loss depending on the subject's action and, if relevant, the actual loss that subsequently occurred that round.

We conducted both studies at a laboratory dedicated to behavioral research. We implemented the experimental software using SoPHIE (Hendriks 2012). Subjects were students at two major public universities in the United States. They volunteered, enrolling through a web-based recruiting system.

Prior to play, subjects read instructions (reprinted in the Appendix A). The monitor then read the instructions aloud and answered any clarification questions. All decisions were made on a computer. In each game, subjects knew that the loss amount was 100 and the cost amount was either 25 or 75, depending on the condition. Forecast information depending on the condition (Neither, Probability, Recommendation, or Both) was made available prior to subjects making a decision of whether to take the risk or take the cost. A random draw from a uniform distribution, consistent with the probability P of loss for that round, determined the outcome of the situation. Draws were independent across subjects, games, and periods and were displayed to the subjects independent of the decision they made. The draw realization and payoff outcome was then displayed to the decision maker. No additional feedback was given between games. At the end of the session, the total tokens a subject earned across all rounds were converted to U.S. dollars at a rate of 1,000 tokens equal to $1, and each subject was paid his or her own earnings in cash. The average earnings per subject, including a $5 show-up fee, were $17. Sessions lasted approximately 45 minutes.

3. Study 1

3.1. Design

Study 1 had a 2 (cost conditions) × 2 (information conditions) × 2 (probability conditions) full factorial design for a total of eight treatments. In the Probability conditions, subjects were shown the loss probability that pertains to that round. In the Recommendation conditions, we described the recommendation as follows:

> Each round, you will be given Advice of whether to Take the Cost or Take the Risk. The Advice has been determined in a way that on average, if you follow the Advice you will earn the most money possible. You are not required to take the advice. Note that the Advice does not guarantee that you will make the most money possible in any given round. It is possible that when the advice is Take the Risk, the Loss does occur. It is also possible that the Advice is to Take the Cost, and the Loss does not occur.

Table 2 summarizes the experimental design and sample sizes. The eight treatments are labeled according to cost and information condition (e.g., "low cost").

3.2. Overall Results

We begin by plotting the proportion of decisions to take the cost over time, for the low cost conditions in Figure 2 and for the high cost conditions in Figure 3.

There are several observations to make. To begin, the figures show that in both "Neither" conditions, decisions are independent of the forecast optimal decisions. This is as expected since subjects in the Neither conditions did not have access to the round-by-round forecast information. Also observe that the majority of actions taken in the Neither conditions are the status quo actions: take the cost in the low cost condition and take the risk in the high cost condition. Next, observe that providing forecasts results in an increase in the proportion of optimal actions in all information conditions. However, different information conditions result in different behaviors in terms of the frequency of observing the optimal action and the variability of behavior.

To further investigate how different information affects behavior, in Figure 4 we plot the proportion of the time the optimal action was chosen in each condition separated by the cost condition, and broken out on whether the optimal action was to take the risk or take the cost. The figure also displays average frequencies and corresponding standard errors based on subject averages.

Table 2. Summary of the Experimental Design (N Is the Sample Size)

	Low cost: $Cost = 25$		High cost: $Cost = 75$	
	Probability not provided	Probability provided	Probability not provided	Probability provided
Recommendation not provided	Low Neither $N = 30$	Low Probability $N = 30$	High Neither $N = 30$	High Probability $N = 30$
Recommendation provided	Low Recommendation $N = 30$	Low Both $N = 30$	High Recommendation $N = 30$	High Both $N = 30$

Figure 2. Low Cost Treatments: Proportion of Subjects Who Take the Cost

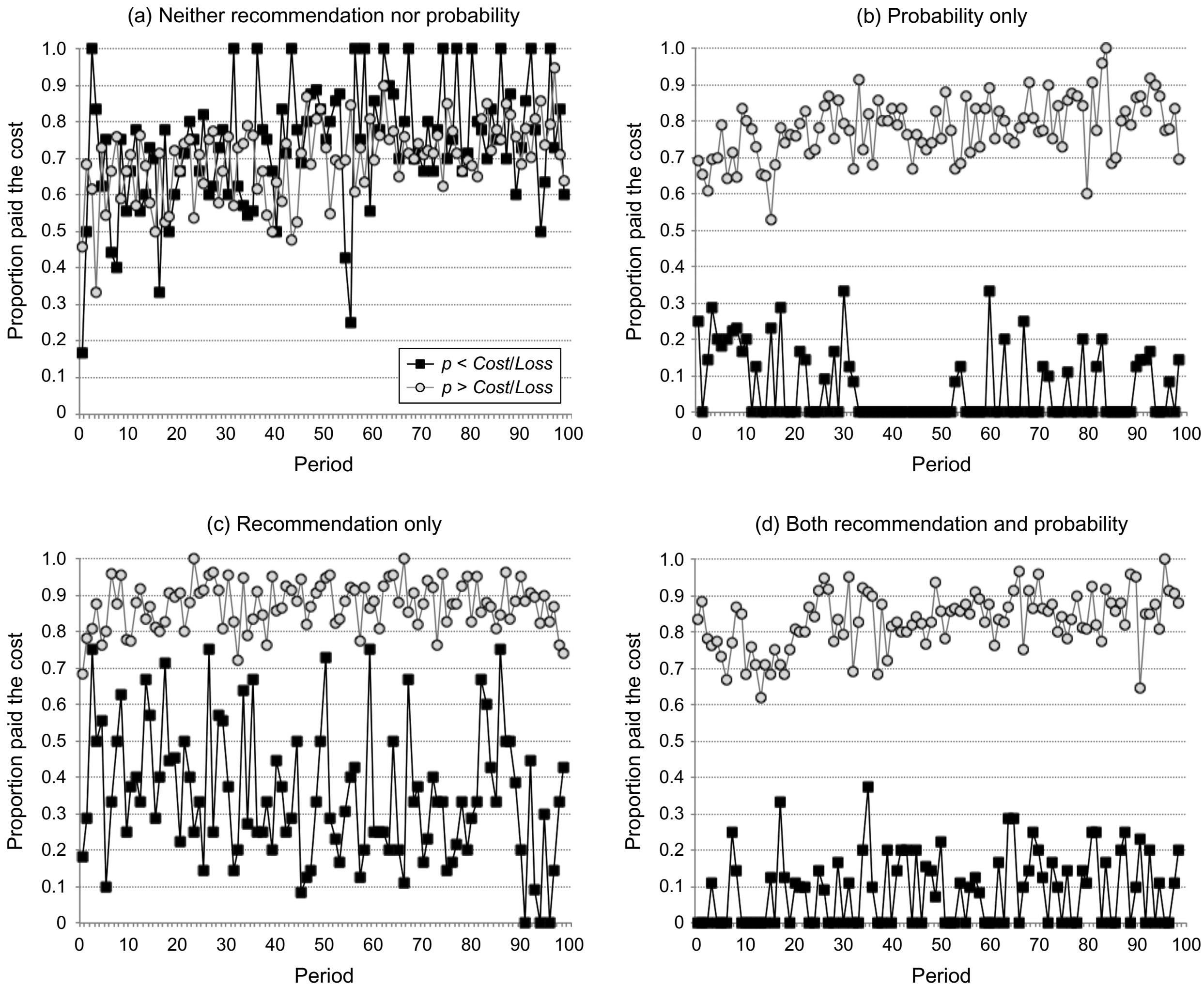
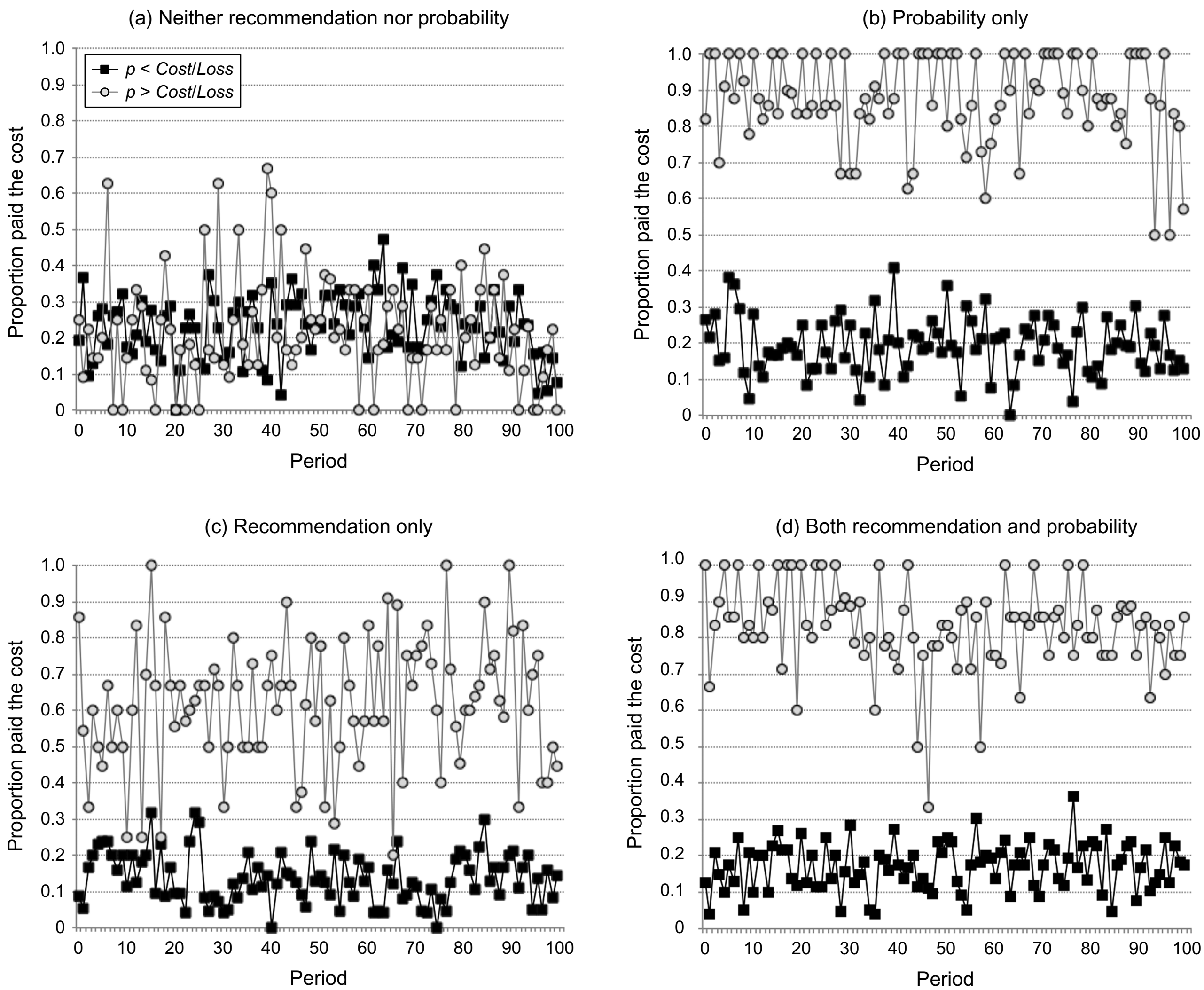


Figure 5 presents results of pairwise rank-sum tests that compare the frequency of taking the optimal action for each pair of conditions in our study. The p-values below 0.10 are highlighted using shaded background.

Figures 4 and 5 display a duality with regard to status quo and siren actions. Specifically, observe the similarity of the frequency of the status quo actions across high and low conditions (take the risk in the high cost condition; take the cost in the low cost condition) in terms of both level and order across information conditions. The same kind of similarity holds for the frequency of siren actions (take the cost in the high cost condition; take the risk in the low cost condition). The pattern of (in)significance in Figure 5 has the same configuration. The duality suggests that the effect of forecast information depends largely on whether the information implies that the decision maker should take the status quo or the siren action, less on whether the status quo (siren) action is to take the cost or risk. In what follows, we use this duality to organize the analysis. Section 3.2 includes formal statistical evidence that the influence of status quo and siren actions is largely independent of the cost variable.

Absent forecast information, subjects take the status quo action with a frequency of 77% and 70% in High Neither and Low Neither conditions, respectively. Getting a recommendation forecast to take the status quo action significantly increases these frequencies to 86% and 87%, although the former increase is only weakly significant. Receiving a probability forecast that implies that the status quo action is optimal increases the frequency of taking the status quo action, but not by as much (81% and 78%) and significantly only in the latter case. Moreover, a direct comparison of recommendation and probability forecasts shows the former to lead to significantly higher compliance rates with the status quo forecast. The conclusion is that recommendation forecasts are more effective at improving compliance with the status quo action than are

Figure 3. High Cost Treatments: Proportion of Subjects Who Take the Cost



probability forecasts. (We compare combination forecast below.)

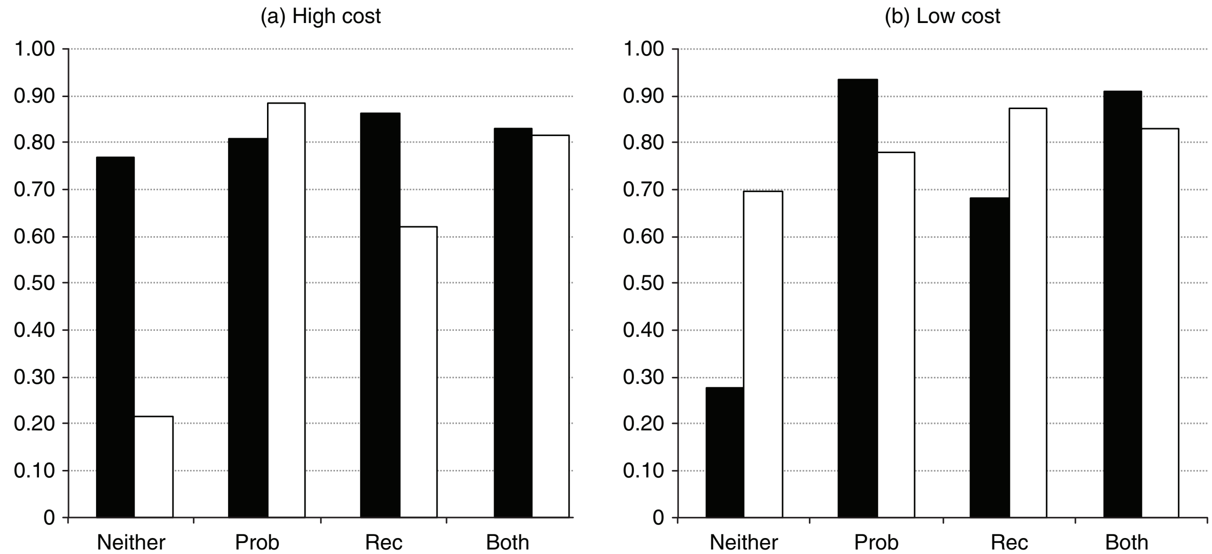
The frequency of the siren action, absent forecast information, is about 22% and 28% in High Neither and Low Neither conditions, respectively. Both recommendation and probability forecasts increase the compliance with the siren action. However, now, the probability forecast exhibits the greatest increase: 88% versus 62% and 94% versus 68% for high and low cost, respectively; both differences are significant. The conclusion is that probability forecasts are more effective at increasing compliance with the siren action than are recommendation forecasts.

Given that recommendation forecasts are best at inducing the compliance with the status quo action and that probability forecasts are best at inducing the compliance with the siren action, we might think that a combination of the two would capture the best of the two effects. This is almost correct. The combination forecast performance is near but not quite as high as the best performances of the individual forecast in both categories: for status quo action, 83% versus 86% and 83% versus 87% for high and low cost, respectively; for siren action, 82% versus 88% and 91% versus 94%, respectively. The compliance frequencies that a combination forecast induces for the status quo (siren) action are not statistically different from those the recommendation (probability) forecast induces.

3.3. Individual-Level Analysis

Figure 6 classifies individuals into five categories based on the frequency with which the individual took the optimal action. Because individuals in the Neither treatment had neither probability nor recommendation information, the optimal action for the Neither treatment subjects was always the status quo; any siren action taken when the underlying round-by-round probability suggests it is optimal is merely coincidental. Overall, the individual-level analysis mirrors the aggregate analysis. Both, probability, and

Figure 4. (Color online) Proportion of the Time the Optimal Action Was Chosen



	(a) Neither	(a) Prob	(a) Rec	(a) Both		(b) Neither	(b) Prob	(b) Rec	(b) Both
■ Risk	0.7690 (0.0343)	0.8092 (0.0215)	0.8617 (0.0267)	0.8316 (0.0274)		0.2766 (0.0461)	0.9360 (0.0153)	0.6827 (0.0722)	0.9108 (0.0408)
□ Cost	0.2168 (0.0348)	0.8843 (0.0336)	0.6200 (0.0520)	0.8172 (0.0554)		0.6971 (0.0426)	0.7807 (0.0283)	0.8735 (0.0342)	0.8312 (0.0289)

Note. Standard errors are in parentheses.

recommendation information are effective in terms of improving compliance with the siren action, but probability information results in significantly more individuals exhibiting high compliance rates (the black bars in Figure 6, panels (a) and (c)) than recommendation by itself. But recommendation is more effective in increasing the number of individuals who comply with the status quo action (the black bars in Figure 6,

Figure 5. (Color online) Pairwise Rank-Sum Tests: Two-Tailed *p*-Values; *N* Is the Number of Subjects

High cost — Optimal action is to take the risk

	Probability	Recommendation	Both
Neither	0.5944	0.0798	0.2573
Probability		0.0212	0.3006
Recommendation			0.3776

High cost — Optimal action is to take the cost

	Probability	Recommendation	Both
Neither	0.0000	0.0000	0.0000
Probability		0.0027	0.8877
Recommendation			0.0003

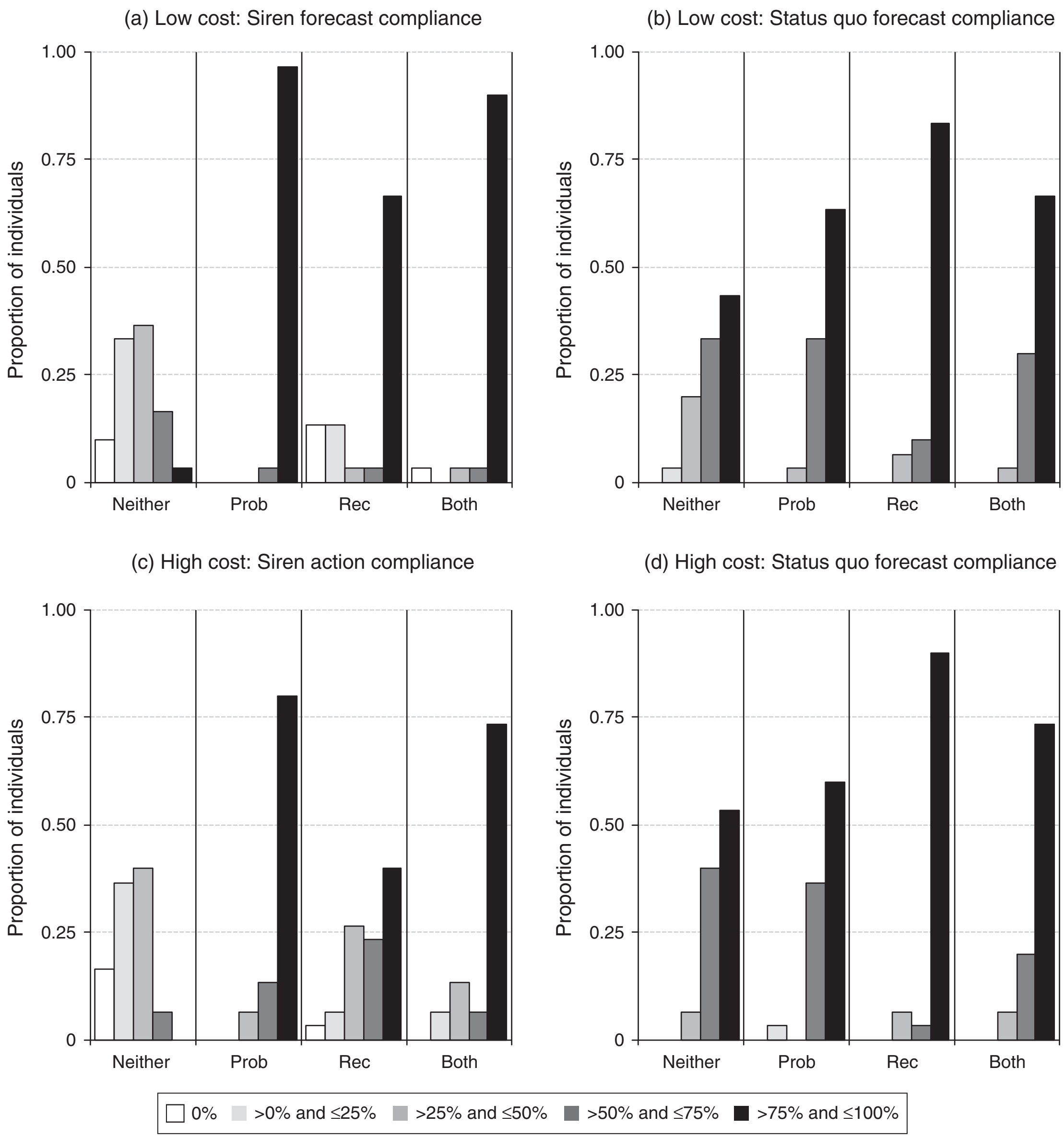
Low cost — Optimal action is to take the risk

	Probability	Recommendation	Both
Neither	0.0000	0.0003	0.0000
Probability		0.0290	0.3227
Recommendation			0.0044

Low cost — Optimal action is to take the cost

	Probability	Recommendation	Both
Neither	0.1983	0.0008	0.0154
Probability		0.0000	0.1331
Recommendation			0.1039

Figure 6. Proportion of Individuals in Compliance with Forecast, by Condition and Optimal Action



panels (b) and (d)). Also, as with the aggregate analysis, compliance rates for the combination forecast are consistently in between those for probability and recommendation forecasts.

To summarize, both recommendation and probability forecasts increase the frequency of compliance with the forecast optimal action relative to the baseline no forecast conditions, although the size of the increase and its significance vary. Recommendations are significantly more effective at inducing status quo action. Probabilities are significantly more effective at inducing the siren action. A forecast that combines probability and recommendation captures most the gains of the separate approaches. These conclusions hold on the aggregate as well as on the individual level.

3.4. The Cry Wolf Effect

What accounts for the differences in performance between recommendation and probability forecasts? The initial hypothesis, that recommendations are more easily applied than are quantitative probabilities, fits well with the data related to status quo forecasts.

But then how can we explain the greater success of probabilities with siren forecasts? We might hypothesize that the reason has to do with risk aversion. The recommendations given to subjects assume they are risk neutral. The hypothesis would be that recommendations of siren actions are followed less often, relative to probability information, because the recommendation overestimates the subject's appetite for risk. The implication is that subjects should learn to follow the recommendation to take the risk *less* frequently (relative to receiving probabilities) in both cost conditions. In fact, they follow the recommendation significantly less often in the high cost conditions but significantly more often in the low cost conditions. So risk aversion is not sufficient to explain the difference in probability and recommendation performance.

An alternative hypothesis is that recommendations for siren actions suffer from a cry wolf effect greater than that associated with probability forecasts. For probability forecasts, we define a forecast as a false alarm if the ex post optimal action is different from the ex ante optimal action implied by the forecasted probability. Table 3 summarizes the estimates of the panel logit models that we fit to explore this hypothesis. All models use random effects variables applied at the subject level to control for individual differences in decision making. We pool the data from the six conditions in which forecast information is provided (Probability, Recommendation, and Both for high and low cost conditions). In doing so, we relabel the cost-loss actions as either the status quo action or the siren action.[4] The second column in Table 3 provides a short description of each independent variable.

The dependent variable in all models is 1 when the siren action is taken and 0 otherwise; so in the high cost condition, the dependent variable is 1 when the decision was to take the cost, and in the low cost condition, the dependent variable is 1 when the decision was to take the risk. The explanatory variable *CostH* is 1 in the high cost condition and 0 in the low cost condition. We include this variable to test for differences attributable to cost level. Variable *Opt* is 1 when the calculated optimal decision given the forecast corresponds to the siren action and 0 otherwise, so in the high cost condition, it is equal to 1 when the forecast implies the optimal decision is to take the cost and 0 otherwise; in the low cost condition, it is equal to 1 when the forecast implies the optimal decision is to take the risk and 0 otherwise. To capture experience effects for taking the status quo and siren actions, we cross the *Period* variable $(1, \ldots, 100)$ with the indicator variable *Opt* and with $(1 - Opt)$.

Model (1) takes a first look at the dynamics of the decision making. Here, as in all models, the *Opt* coefficient is positive and significant, thereby confirming that, overall, forecast information improves decision quality. Also, as in all, the *CostH* variable is not significant, indicating no difference in the pattern of status quo/siren decision making across cost conditions. On aggregate, experience leads subjects to learn to take the optimal action more often. The coefficient for $Period \times Opt$ is positive but not significant, and for $Period \times (1 - Opt)$, it is negative and significant. The pattern is repeated in the other three models, and in Models (3) and (4), both coefficients become significant. So on aggregate, and with experience, subjects learn to take the optimal action, be it the status quo or siren action, more often with experience.

Table 3. Effect of Information on Choice Dynamics

		Dependent variable = 1 if siren action chosen			
Explanatory variable	Description	(1)	(2)	(3)	(4)
CostH	Equals 1 when *cost* = 75; 0 otherwise	−0.1246	−0.1238	−0.1116	−0.1113
		(0.2443)	(0.2556)	(0.2489)	(0.2433)
Opt	Equals 1 when the siren action is optimal; 0 otherwise	3.836***	3.8370***	3.8440***	3.7955***
		(0.1135)	(0.1136)	(0.1141)	(0.1142)
$Period \times Opt$	The period number (1 to 100) × *Opt*	0.0018	0.0029	0.0230***	0.0248***
		(0.0016)	(0.0039)	(0.0059)	(0.0059)
$Period \times (1 - Opt)$	The period number (1 to 100) and × (1 − *Opt*)	−0.0044***	−0.0033	−0.0096**	−0.0095**
		(0.0009)	(0.0114)	(0.0040)	(0.0040)
Wolf	Cumulative number of false alarms in current period		−0.0034	0.0164	0.0161
			(0.0114)	(0.1220)	(0.0121)
$Wolf \times Opt$	*Wolf* interacted with the siren forecast			−0.0829***	−0.0700***
				(0.0176)	(0.0178)
$Wolf \times Rec \times Opt$	*Wolf* interacted with the siren forecast in Rec condition				−0.0401***
					(0.0064)
Constant		−1.9002***	−1.9016***	−1.9189***	−1.8986***
		(0.1791)	(1,794)	(0.1828)	(0.1786)
Log likelihood		−6,932.0	−6,932.0	−6,920.8	−6,900.8
Observations				18,000	
Groups				1,809	

Notes. Random effects logit. Standard errors are in parentheses.
*p-value < 0.10; **p-value < 0.05; ***p-value < 0.01 (two-tailed).

For Model (2) we define the variable *Wolf* to be the total number of previous forecasts that turned out to be incorrect ex post—either the forecasted probability was greater than *Cost/Loss* but the loss did not occur or the forecasted probability was smaller than *Cost/Loss* but the loss did occur. So the *Wolf* variable is simply the total number of false alarms observed during the session through the current period. In Model (2) *Wolf* is negative but not significant, and adding it does not change either the sign or the significance of any of the other variables.

For the last two models we cross the variable *Wolf* with an indicator variable, *Opt*, which, as we explained above, is 1 when the optimal action is the siren action (take the risk in the low cost condition or take the cost in the high cost condition) and 0 otherwise. In Model (3) we add the $Wolf \times Opt$ variable to the explanatory variables in Model (2). $Wolf \times Opt$ is negative and highly significant, indicating that subjects become less likely to take the siren action the more false alarms they observed in previous periods.

For Model (4) we add a variable, $Wolf \times Opt \times Rec$, that interacts the $Wolf \times Opt$ variable with the indicator variable for recommendation only (without probability information). The coefficient to $Wolf \times Opt$ is still significantly negative, indicating that there is a cry wolf effect associated with providing probability information (with or without recommendation information). The coefficient to $Wolf \times Opt \times Rec$ is negative and significant, indicating that in treatments in which the recommendation is provided without probability, when the recommendation is to take the siren action, observing more false alarms causes subjects to become less likely to comply with future siren forecasts.

We ran a number of robustness checks on the results reported in Table 3. The results are summarized in the more extensive model estimates presented in Table B.1 in Appendix B. The latter model extends Model (4) of Table 3 by adding controls for forecast conditions, more extensive checks for differences across high and low cost conditions, and a check for a cry wolf effect in the combined probability–recommendation condition. All of the results reported in this section continue to hold with respect to estimated coefficient signs, size, and statistical significance levels. The estimated cry wolf effect for the combined probability–recommendation condition is not significantly different from that for the probability-alone condition. Also, the cry wolf effect is somewhat, and significantly, higher in the high cost conditions.

An additional robustness check, presented in Table B.2 in Appendix B, estimates the four models in Table 3, replacing the *Wolf* variable by the *Wolf/Period*—the average number of false alarms observed per period. Results in terms of the signs and significance levels of the variables that measure the cry wolf effect with respect to the compliance with the siren action are not affected.

To summarize, all of the forecast information forms we provided exhibit a cry wolf effect, which becomes larger when the recommendation is provided without probability information. The statistical analysis in Table 3 shows that false alarms (the *Wolf* variable in the analysis) detract from compliance with a siren forecast but not a status quo forecast. The intuition behind this result is that the loss of credibility from false alarms leads users to ignore the forecast, falling back on what they would do if no forecast were available. From the Neither treatments, what most would do is take the status quo action, implying the asymmetric influence on siren and status quo forecasts. Recommendation forecasts lead to a larger cry wolf effect than probability forecasts. The larger effect is consistent with the hypothesis that cry wolf effects are on a continuum, the size of the cry wolf effect from false alarms depending on the perceived gap between the stated and actual uncertainty.

4. Study 2

Study 2 was designed as a robustness check on Study 1 results showing recommendations that prove ex post incorrect suffer a higher loss of forecast credibility than the analogous probability forecast. Specifically, Study 1 instructions stated that the recommendation "has been determined in a way that on average, if you follow the Advice you will earn the most money possible," but the instructions provided no guidance on the method used to construct the recommendation. In Study 2, we study the hypothesis that a recommendation forecast would be less prone to the cry wolf effect, and therefore more effective, if users are informed of how the recommendation is formulated from the quantitative risk measure. If the hypothesis proves correct, then explaining that risk was taken into account can substitute for a forecast statement of the quantitative risk.

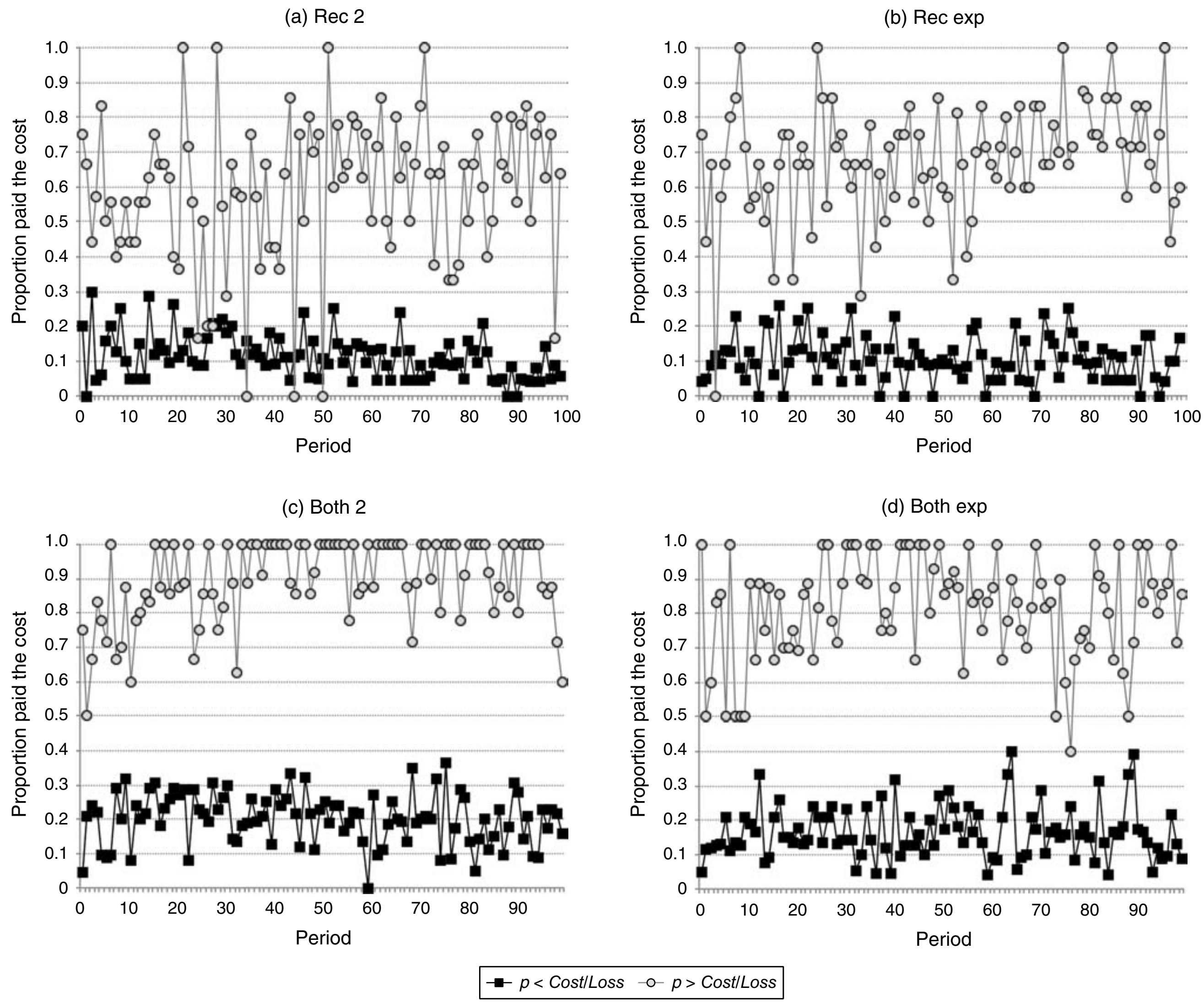
4.1. Design

Study 2 instructions given to recommendation forecast users add an explanation of how cost, loss, and probability of loss are used to derive the recommendation:

> The Advice is based on comparing the Cost to the expected value of the Loss. If the expected value of the Loss is greater than the Cost, then the Advice will be to Take the Cost. If the expected value of the Loss is less than the Cost, then the Advice will be to take the Risk. In case of a tie (rare), the advice is determined at random.
>
> $$\text{The expected value of the Loss} = \text{Loss Probability} \times 100$$
>
> It is the average loss we would expect over the long run (over many rounds) for a certain Loss Probability.

Management Science, 2018, vol. 64, no. 3, pp. 1440–1457, © 2017 INFORMS

Figure 7. Study 2 Treatments: The Proportion of Subjects Who Take the Cost



Study 2 included four treatments, all in the high cost condition. The treatments were as follows:

1. *Recommendation only* (*Rec* 2), which is a replication of the recommendation treatment of Study 1 ($N = 29$).
2. *Recommendation with explanation* (*Rec exp*) included the explanation above ($N = 29$).
3. *Recommendation and probability* (*Both* 2), which is a replication of the both treatment of Study 1 ($N = 30$).
4. *Recommendation and probability with explanation* (*Both exp*) also provided probability information ($N = 30$).

We conducted Study 2 at a university with a somewhat different student demographic than the university in which we conducted Study 1. Prior to running this study, we hypothesized the nature of the different demographic would lead to more forecast adherence (analyzed in Section 4.4).

4.2. Overall Results

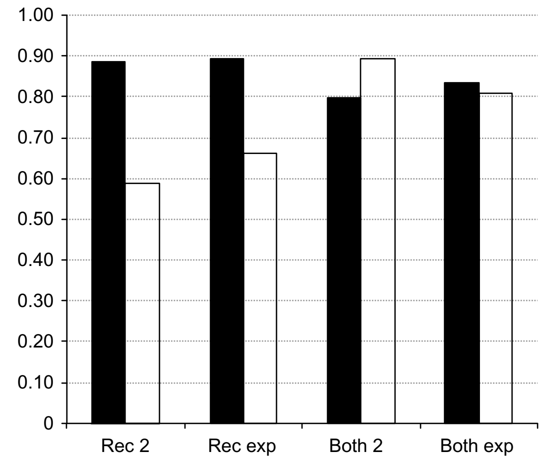
Figure 7 plots the proportion of decisions that take the cost over time for the four new treatments.

To further investigate how different information affects behavior, in Figure 8 we plot the proportion of the time the optimal action was chosen in the four Study 2 treatments, separated by the cost condition, and conditioning on whether the optimal action was to take the risk or take the cost. The figure also displays average frequencies and corresponding standard errors based on subject averages.

Figure 9 presents results of pairwise rank-sum tests that compare the frequency of taking the optimal action for each pair of conditions in Study 2. The p-values below 0.10 are highlighted using shaded background.

It is clear from the above analysis that the main driver for improving the frequency of taking the

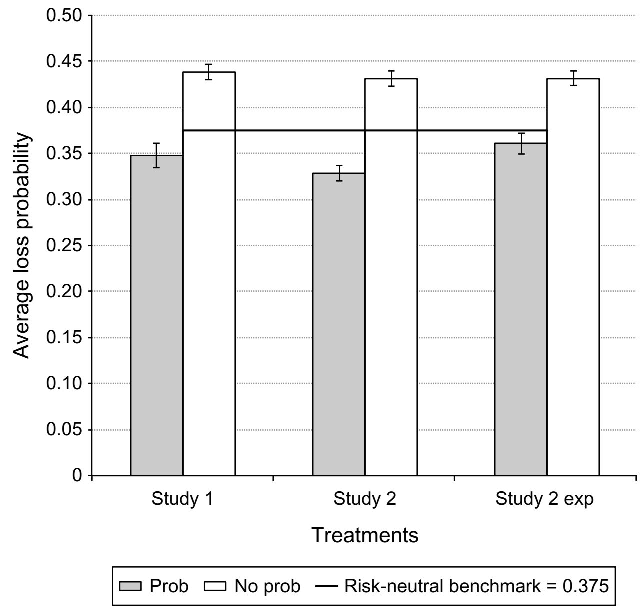
Figure 8. (Color online) Proportion of the Time the Optimal Action Was Chosen



■ Risk	0.8844 (0.0207)	0.8934 (0.0275)	0.7968 (0.0245)	0.8355 (0.0212)
□ Cost	0.5877 (0.0596)	0.6632 (0.0494)	0.8920 (0.0215)	0.8093 (0.0432)

Note. Standard errors are in parentheses.

Figure 10. Average Loss Probabilities for the "Take the Risk" Decisions



optimal action is the presence of probability information. When added to a recommendation (with or without the explanation), showing the probability increases the frequency of taking the optimal action when this action is the siren action (take the cost) and decreases this frequency when the optimal action is the status quo action (take the risk). This is the same pattern that we observed in Study 1 for the siren action. For the status quo action, the decrease in following the optimal policy, observed when the probability is added to a recommendation, is statistically significant in Study 2 both when the explanation is given and when it is not. In Study 1 we also observed a decrease, but it was not statistically significant. The small decreases in the frequency of complying with the siren action, in Study 2, can be rationalized by some degree of risk aversion. A risk-averse individual would have a lower cutoff probability of a loss for switching from taking the cost to taking the risk than would a risk-neutral individual. This implies that for a risk-averse individual, the average loss probability at which the individual decided to take the risk would be lower than the same average loss probability for a risk-neutral individual. If loss probability is not shown, risk-averse individuals may still take the risk less frequently; however, there is no reason to expect any systematic effect on the average loss probability since these probabilities are not visible to them. In Figure 10 we plot average loss probabilities we observe in our studies in treatments with known loss probability (Prob) and analogous treatments with unknown loss probability (No prob). The figure also shows the risk-neutral benchmark.

The average loss probability for take the risk decisions is significantly lower in Prob treatments than in analogous No prob treatments ($p < 0.001$ for all three comparisons). More germane to the noted risk hypothesis, the average loss probability in No prob treatments is above the risk-neutral benchmark of 0.375. And, by contrast, the average loss probability in Prob treatments is below the risk-neutral benchmark, significantly so for treatments without explanation ($p = 0.0549$ in Study 1 and $p < 0.001$ in Study 2) and not significantly for the treatments with explanation

Figure 9. (Color online) Pairwise Rank-Sum Tests: Two-Tailed p-Values; N Is the Number of Subjects

Optimal action is to take the risk

	Both 2	Rec exp	Rec 2
Both exp	0.5596	0.0009	0.0807
Both 2		0.0111	0.0046
Rec exp			0.2339

Optimal action is to take the cost

	Both 2	Rec exp	Rec 2
Both exp	0.2772	0.0000	0.0017
Both 2		0.0110	0.0000
Rec exp			0.4413

($p = 0.212$). So there is some evidence that risk aversion plays a role, although the role is rather small.

Providing the explanation for the recommendation seems to slightly increase the frequency of taking the optimal siren action when probability is not available, but because the increase is not significant, we conclude that the explanation has no clear effect.

4.3. The Cry Wolf Effect

Table 4 summarizes the Study 2 estimates of the logit models for the same Models (1), (2), and (3) used in Study 1 (see Table 3). We omit the description column, as it is largely redundant. In Model (4) we measure the effect of probability information and recommendation explanation. Data from all four Study 2 treatments are used to obtain the estimated coefficients.

The estimates for Models (1), (2), and (3) are nearly identical in their magnitude and significance to the analogous models in Study 1. As before, the *Wolf* variable is not significant, but the *Wolf* × *Opt* variable in Model (3), that measures the cry wolf effect when the siren action is optimal, is negative and significant, indicating that on the whole, the cry wolf effect is present and degrades performance.

In Model (4) we measure the relative influence of our treatments on the cry wolf effect by including interaction variables with probability information (*Prob*) and the explanation of the recommendation (*Exp*). Probability information improves performance by partially mitigating the cry wolf effect, as does the explanation. The improvement from probability is larger ($\chi^2 = 3.47$, $p = 0.0624$ (one-sided)). The interaction variable *Wolf* × *Prob* × *Exp* is negative and significant, and the magnitude is the same as the *Wolf* × *Opt* × *Prob* variable ($\chi^2 = 0.01$, $p = 0.9157$ (one-sided)). This indicates that there is no additional benefit from providing both the explanation of the recommendation and the probability. In fact, including both pieces of information slightly increases the cry wolf effect.

Table 4. Effect of Information on Choice Dynamics

	Dependent variable = 1 if siren action chosen			
Explanatory variable	(1)	(2)	(3)	(4)
Opt	3.0585***	3.0644***	3.0623***	3.0043***
	(0.1191)	(0.1192)	(0.1193)	(0.1199)
Period × *Opt*	0.0073	0.0126***	0.0301***	0.0250***
	(0.0017)	(0.0048)	(0.0066)	(0.0068)
Period × (1 − *Opt*)	−0.0028***	0.0025	−0.0050	−0.0053
	(0.0011)	(0.0046)	(0.0158)	(0.0050)
Wolf		−0.0170	0.00072	0.0082
		(0.0145)	(0.0205)	(0.0157)
Wolf × *Opt*			−0.0804***	−0.0929***
			(0.0204)	(0.0211)
Wolf × *Opt* ×*Prob*				0.0593***
				(0.0114)
Wolf × *Opt* ×*Exp*				0.0370***
				(0.0091)
Wolf × *Opt* ×*Prob* × *Exp*				−0.0583***
				(0.0151)
Constant	−2.0735***	−2.0756***	−2.0767***	−2.0537***
	(0.1375)	(0.01378)	(0.1377)	(0.1346)
Log likelihood	−4,659.7	−4,659.0	−4,651.3	−4,632.7
Observations			11,800	
Groups			118	

Notes. Random effect logit. Standard errors are in parentheses.
*p-value < 0.10; **p-value < 0.05; ***p-value< 0.01 (two-tailed).

4.4. Subject Pool Effect

For completeness, we tested whether our results are sensitive to the subject pool. Study 1 was conducted at a large public university in the rural part of the northeastern United States; subjects were mostly undergraduates, predominantly business or engineering majors, and foreign students constituted about 40% of that subject pool. Study 2 was conducted at a medium-sized university located in a large metropolitan area in Texas; subjects were predominantly foreign graduate students studying business or engineering. We do not find any subject pool effect when comparing the frequency of taking the optimal action (siren or status quo) in Rec and Rec 2 or Both and Both 2 (all p-values > 0.1).

We also estimated a logit model that included the four treatments: Rec, Rec 2, Both, and Both 2, with the first four independent variables the same as in Model (2) (see Table 4) and additional independent variables *Pool* and *Wolf* × *Opt* × *Pool*, where *Pool* is the indicator variable for treatments Rec 2 and Both 2. Neither variable was significant ($p = 0.707$ for *Pool* and $p = 0.768$ for *Wolf* × *Opt* × *Pool*), indicating that the cry wolf effect we reported is stable for two substantially different subject pools.

5. Summary and Discussion

Decisions under risk are increasingly made with the aid of expert forecasts. We report on results of a laboratory experiment designed to compare the relative effectiveness of two ways of providing forecast information: the probability of the uncertain event versus an unequivocal recommendation. Decision theory says that probabilities are sufficient to make the optimal decision. The potential value of an unequivocal recommendation (of action) is that it sidesteps the errors boundedly rational decision makers are prone to make, in applying quantitative measures such as probabilities, in solving decision problems. The potential downside is that an unequivocal recommendation that proves incorrect in hindsight might induce a cry wolf effect, reducing the decision maker's compliance with future recommendations.

We studied these issues in a laboratory cost-loss game in two studies. In Study 1, subjects played in four forecast information conditions: Neither (no forecast information), Probability, Recommendation, and Both (probability and recommendation combined) and two cost settings (high and low). In Study 2 we conducted two treatments in which we provided a detailed explanation of how the recommendation is determined, and we compared the effect of these conditions to the recommendation only as well as the recommendation and probability conditions.

Regarding the efficacy of forecast formats, our findings from Study 1 are as follows: First, with a prior probability to go on and no period-by-period forecast (the Neither treatments), subjects are about three times more likely to take the expected value optimal action—termed the status quo action—than to take the alternative siren action. Second, the influence of forecast information exhibits a strong duality in that, for a given forecast, the frequency with which an action is taken depends on whether it is status quo or siren, independent of whether the status quo (siren) action is to take the cost or take the risk. Third, both probability and recommendation forecasts increase the frequency of expected value optimal actions, although the pattern of effectiveness differs. Recommendation forecasts are more effective at inducing status quo action when it is optimal. Probabilities are more effective at inducing siren action when it is optimal. A combination forecast, providing both a recommendation and probability, captures most of the benefits of the individual approaches.

In Study 2 we provide a robustness check of those conclusions. We find that explaining how recommendation is determined does not significantly improve the frequency of taking either the status quo or the siren action, although the explanation does have a small mitigating effect on the cry wolf effect. We confirmed the finding that adding probability information on top of the recommendation increases the frequency of taking siren action when it is optimal. It also somewhat decreases the frequency of taking the status quo action when it is optimal. These results hold regardless of whether the recommendation was accompanied by an explanation.

Regarding the reasons for the differences in efficacy we identified, the differences in performance between probability and recommendation forecasts cannot be fully explained by risk aversion (because the frequency of talking the cost, when it is optimal, should not be affected by risk aversion), although we found evidence that risk aversion may play some role. The pros and cons of recommendation forecasts can explain much more of the difference: recommendations are more straightforward than probabilities to turn into optimal decisions. At the same time, they suffer from a cry wolf effect when the unequivocal recommendation turns out to be wrong in hindsight, reducing the decision maker's adherence to future recommendations. What our data show is that the cry wolf effect extends to probability forecasts as well, although it is higher for recommendations. Both probability and recommendations reduce compliance with siren forecasts, but more so for recommendations.

Our work has implications for designing forecast guidance and warning systems that are meant to convey expert forecast information to nonexperts who use the information to make decisions under risk. Perhaps most noteworthy, we find that providing probability information is most important in situations in which the recommendation is a siren action—an action that is contrary to the status quo action that would prevail absent a forecast. At the same time, our findings suggest that there is room for improvement since all forecasts we studied suffer from a significant cry wolf effect. Future work on forecast designs seeking to mitigate the cry wolf effect could have a large impact on the efficacy of expert forecast systems.

Acknowledgments

The authors thank Yang Zhang for conducting the sessions at Penn State, and Mrunal Hadke for conducting the sessions at the University of Texas at Dallas. The authors also thank the Laboratory for Behavioral Operations and Economics at the Naveen Jindal School of Management, the University of Texas at Dallas, for providing the facilities and financial support. The authors gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft through the DFG-Research Group "Design & Behavior" and its members for useful comments.

Appendix A

Figure A.1 shows the instructions that were given to the subjects for the high cost Both exp treatment. In this treatment, the subjects received both probability and advice with a cost of 75 tokens and an explanation for how the advice was determined. When the subjects were given only probability, the paragraphs detailing the advice were removed. When the subjects were given only advice, the details of the probability were removed. When the subjects were given advice without explanation, the explanation paragraph was removed. When the subjects were given neither advice nor probability, they were provided only with the loss probability across all rounds. For the conditions in which the cost was 25 tokens, each mention of "75" was changed to a "25" in the instructions.

Figure A.1. Instructions Given to Subjects for the High Cost Both Exp Treatment

INSTRUCTIONS

General. The purpose of this session is to study how people make decisions in a particular situation. If you have any questions, feel free to raise your hand and a monitor will assist you. From now until the end of the session, unauthorized communication of any nature with other subjects is prohibited.

During the session you will play 100 rounds of a game from which you can earn tokens. At the end of all the games, the tokens you earn will be converted into dollars; the more tokens you earn, the more money you will make.

Description of the game. At the beginning of each round of the game, you are given a credit of 150 tokens. You must then decide whether to **Take the risk** or **Take the cost**. If your decision is to Take the risk, there is some probability of incurring a **Loss** of 100 tokens. If your decision is to Take the Cost, you incur a **Cost** of 75 tokens for certain.

Your Profit depends on your decision and on whether the Loss occurs.

If you Take the risk, then either

Your Profit = 150-100 = 50 tokens <u>if the Loss occurs</u>

or

Your Profit = 150 tokens if the Loss does not occur.

If you Take the Cost, then

Your Profit = 150-75 = 75 tokens regardless of whether the Loss occurs or not.

The Probability of Loss varies from round-to-round. To determine whether the Loss actually occurs in a round, the computer will generate a random number between 0 and 1, with each number in this range equally likely. If the random number is below or equal to the Loss Probability for that round, the Loss occurs; if it is above the Loss probability, the Loss does not occur.

For example, suppose the Loss Probability for the round is 0.60. If the random number comes out to be .65, the Loss does not occur. If the random number comes out to be 0.4, the Loss occurs.

Information to help you decide. While the Loss Probability varies from round-to-round, the average Loss Probability across all rounds is 0.50.

Probability

Each round you will be given the Loss Probability that pertains to that round.

Recommendation

Each round you will be given Advice of whether to Take the Cost or Take the risk. The Advice has been determined in a way that on average, if you follow the Advice you will earn the most money possible. You are not required to follow the advice.

Note that the Advice does not guarantee that you will make the most money possible in any given round. It is possible that when the advice is Take the risk, the Loss does occur. It is also possible that the Advice is to Take the Cost, and the Loss does not occur.

Explanation

The Advice is based on comparing the Cost to the expected value of the Loss. If the expected value of the Loss is greater than the Cost, then the Advice will be is to Take the Cost. If the expected value of the Loss is less than the Cost, then the Advice will be is to Risk. In case of a tie (rare), the advice is determined at random.

The expected value of the Loss = Loss Probability × 100

It is the average loss we would expect over the long run (over many rounds) for a certain Loss Probability.

Payment. At the end of the session the actual earnings from all 100 rounds will be converted to US dollars at the rate of 1000 tokens for 1 US dollar. These profits will be displayed on your screen and paid to you in cash, along with the $5 participation fee, at the end of the session.

Appendix B

We present two additional regressions that provide a slightly more nuanced analysis of the dynamics that we presented in Table 3, and they demonstrate that our conclusions are quite robust. Model (1) is reprinted here for convenience—it is the same as Model (4) in Table 3. In Model (2) we add variables for the recommendation only treatments (*Rec*) and the Both Treatments (*Rec* × *Prob*). The *Rec* variable is negative and significant, indicating that the compliance with the siren action is lower in the recommendation only treatment than in the probability only treatment (the baseline in these regressions). The *Rec* × *Prob* variable is positive and significant, indicating that adding probability to recommendation increases compliance with the siren action. Variables that interact *Rec* and (*Rec* × *Prob*) with *CostH* are not significant, confirming the behavioral symmetry in regard to complying with the siren action, in the high cost and low cost conditions. In Model (3) we test the extent to which there are differences in dynamics in the high cost condition. The variable (*Wolf* × *Opt*) × *CostH* in Model (3) is negative and significant, while the original *Wolf* × *Opt* variable also remains negative and significant, although the magnitude becomes smaller. This indicates that the cry wolf effect with respect to compliance with the siren action exists in both cost conditions, but it is a bit stronger in the high than in the low cost condition.

Table B.2 presents the same set of models as in Table 3, but the *Wolf* variables are normalized by dividing the cumulative number of false alarms by the number of periods. This change does not result in any fundamental differences in either the signs or the significance levels of the important variables. One difference is that in Model (3), variable *Wolf/Period* is positive and significant. The jump in the size of *Wolf/Period* from Model (2) may indicate some collinearity between *Wolf/Period* and other variables in the model—a potential reason to prefer the Table 3 specification.

Table B.1. Robustness Check for the Effect of Treatment Variables on Choice Dynamics

	Dependent variable = 1 if siren action chosen		
Explanatory variable	Model (1)	Model (2)	Model (3)
CostH	−0.111	−0.286	−0.149
	(0.243)	(0.399)	(0.409)
Opt	3.795***	3.802***	3.802***
	(0.114)	(0.114)	(0.115)
Period × *Opt*	0.0248***	0.0250***	0.0263***
	(0.006)	(0.006)	(0.006)
Period × (1 − *Opt*)	−0.00948*	−0.00962*	−0.00908*
	(0.004)	(0.004)	(0.004)
Wolf	0.0161	0.0165	0.0148
	(0.012)	(0.012)	(0.012)
Wolf × *Opt*	−0.0700***	−0.0725***	−0.0471*
	(0.018)	(0.018)	(0.020)
(*Wolf* × *Opt*) × *Rec*	−0.0401***	−0.0370***	−0.0249*
	(0.006)	(0.006)	(0.012)
Rec		−1.479***	−1.683***
		(0.404)	(0.417)
Rec × *CostH*		0.534	0.845
		(0.567)	(0.585)
Rec×*Prob*		1.016*	1.174**
		(0.405)	(0.418)
(*Rec* × *Prob*) × *CostH*		−0.532	−0.783
		(0.569)	(0.587)
(*Wolf* × *Opt*) × *CostH*			−0.0500***
			(0.009)
Opt × *CostH*			−0.0145
			(0.014)
(*Wolf* × *Opt*) × (*Rec* × *Prob*)			0.0038
			(0.009)
Constant	−1.899***	−1.256***	−1.321***
	(0.179)	(0.285)	(0.292)
Log likelihood	−6,900.8	−6,891.5	−6,855.2
Observations (groups)		18,000	
		180	

Note. Standard errors are in parentheses.
*p-value < 0.10; **p-value < 0.05; ***p-value < 0.01 (two-tailed).

Table B.2. Robustness Check for the Using the Normalized Wolf Variable

	Dependent variable = 1 if siren action chosen			
Explanatory variable	Model (1)	Model (2)	Model (3)	Model (4)
CostH	−0.125	−0.126	−0.118	−0.124
	(0.244)	(0.244)	(0.248)	(0.241)
Opt	3.836***	3.836***	4.750***	4.725***
	(0.113)	(0.113)	(0.209)	(0.210)
Period × *Opt*	0.0019	0.0018	0.0030	0.0034*
	(0.002)	(0.002)	(0.002)	(0.002)
Period × (1 − *Opt*)	−0.00437***	−0.00442***	−0.00468***	−0.00468***
	(0.0009)	(0.0009)	(0.0009)	(0.0009)
Wolf/Period		0.120	0.756*	0.785*
		(0.288)	(0.314)	(0.314)
Wolf/Period × *Opt*			−3.175***	−1.592*
			(0.579)	(0.624)
Wolf/Period × *Opt* × *Rec*				−3.881***
				(0.391)
Constant	−1.900***	−1.933***	−2.133***	0.912***
	(0.179)	(0.196)	(0.203)	(0.121)
Log likelihood	−6,932.0	−6,931.9	−6,916.8	−6,867.4
Observations (groups)	18,000			
	180			

Note. Standard errors are in parentheses.
*p-value < 0.10; **p-value < 0.05; ***p-value < 0.01 (two-tailed).

Nevertheless, the estimates for the variables of interest, *Wolf/Period* × *Opt* and *Wolf/Period* × *Opt* × *Rec*, remain negative and significant, as in the corresponding models in Table 3.

Endnotes

[1] Rottenstreich and Kivetz (2006) argue that there are two kinds of real-world decisions: ones in which people rely on estimating probabilities and others in which people instead rely on intuition and other nonprobabilistic cues.

[2] If the subject is risk averse, then she should weigh the cost against the expected utility of the loss. But given the modest stakes of our experiment and the amount of repetition, the optimal rule based on expected losses is likely a good approximation of optimal play for all but the very risk-averse players. Most important, the major conclusions we draw are robust to the assumption that players are risk averse (see Section 3.2).

[3] A potential pitfall is heterogeneity of costs and potential losses. In this case, a recommendation that is right for one forecast user may not be right for another. We do not deal with this issue here.

[4] Recall that the status quo action is the action that is optimal absent a round-by-round forecast (take the cost/risk in the low cost/high cost condition). The siren action is the alternative to the status quo action. It is only optimal if the round-by-round forecast implies such.

References

Arkes HR, Dawes RM, Christensen C (1986) Factors influencing the use of a decision rule in a probabilistic task. *Organ. Behav. Human Decision Processes* 37(1):93–110.

Bliss JP, Gilson RD, Deaton JE (1995) Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics* 38(11):2300–2312.

Breznitz S (1984) *Cry Wolf: The Psychology of False Alarms* (Lawrence Erlbaum Associates, Englewood Hills, NJ).

Budescu DV, Wallsten TS (1987) Subjective estimation of precise and vague uncertainties. Wright G, Ayton P, eds. *Judgmental Forecasting* (John Wiley & Sons, New York), 63–82.

Budescu DV, Por HH, Broomell S (2012) Effective communication of uncertainty in the IPCC reports. *Climate Change* 113(2): 181–200.

Craft DL, Wein LM, Wilkins AH (2005) Analyzing bioterror response logistics: The case of anthrax. *Management Sci.* 51(5):679–694.

Erev I, Bornstein G, Wallsten TS (1993) The negative effect of probability assessment on decision quality. *Organ. Behav. Human Decision Processes* 55(1):78–94.

Fox CR, Tversky A (1998) A belief-based account of decision under uncertainty *Management Sci.* 44(7):879–895.

Green LV, Kolesar PJ (2004) Improving emergency responsiveness with management science. *Management Sci.* 50(8):1001–1014.

Harris AJ, Corner A (2011) Communicating environmental risks: Clarifying the severity effect in interpretations of verbal probability expressions. *J. Experiment. Psych.: Learning, Memory, Cognition* 37(6):1571–1578.

Hendriks A (2012) SoPHIE—Software platform for human interaction experiments. Working paper, University of Osnabrück, Osnabrück, Germany.

Kahneman D (2003) Maps of bounded rationality: Psychology for behavioral economics. *Amer. Econom. Rev.* 93(4):1449–1475.

Karelitz TM, Budescu DV (2004) You say "probable" and I say "likely": Improving interpersonal communication with verbal probability phrases. *J. Experiment. Psych.: Appl.* 10(1):25–41.

Meyer J, Bitan Y (2002) Why better operators receive worse warnings. *Human Factors* 44(3):343–353.

Papastavrou JD, Lehto MR (1996) Improving the effectiveness of warnings by increasing the appropriateness of their information content: Some hypotheses about human compliance. *Safety Sci.* 21(3):175–189.

Pinker EJ (2007) An analysis of short-term responses to threats of terrorism. *Management Sci.* 53(6):865–880.

Rottenstreich Y, Kivetz R (2006) On decision making without likelihood judgment. *Organ. Behav. Human Decision Processes* 101(1): 74–88.

Roulston MS, Smith LA (2004) The boy who cried wolf revisited: The impact of false alarm intolerance on cost-loss scenarios. *Weather Forecasting* 19(2):391–397.

Roulston MS, Bolton GE, Kleit AN, Sears-Collins AL (2006) A laboratory study of the benefits of including uncertainty information in weather forecasts. *Weather Forecasting* 21(1):116–122.

Silver N (2012) *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* (Penguin Press, London).

U.S. National Research Council (2006) *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts* (National Academies Press, Washington, DC).

Wallsten TS, Budescu DV, Zwick R, Kemp SM (1993) Preferences and reasons for communicating probabilistic information in numerical or verbal terms. *Bull. Psychonomic Soc.* 31(2):135–138.

Wallsten TS, Budescu DV, Rapoport A, Zwick R, Forsyth B (1986) Measuring the vague meanings of probability terms. *J. Experiment. Psych.* 115(4):348–365.

Weber EU (1994) From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psych. Bull.* 115(2):228–242.

Weber EU, Hilton DJ (1990) Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *J. Experiment. Psych.: Human Perception Performance* 16(4): 781–789.

Yaniv I, Foster DP (1995) Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *J. Experiment. Psych.* 124(4):424–432.

Yaniv I, Foster DP (1997) Precision and accuracy of judgmental estimation. *J. Behav. Decision Making* 10(1):21–32.