Aggregation Fallacy: An Explanation for Systematic Forecast Bias

Karl Schuhmacher
Emory University
kschuhm@emory.edu (ORCID: 0000-0001-9652-3057)

Jacob T. Zureich
Lehigh University
jaz423@lehigh.edu (ORCID: 0000-0003-1529-0920)

Michael Burkert
University of Fribourg
michael.burkert@unifr.ch (ORCID: 0000-0002-5309-646X)

September 2025

Running Head: Aggregation Fallacy: An Explanation for Systematic Forecast Bias

The authors have no conflicts of interest related to this research.

We appreciate helpful comments and suggestions from Eddy Cardinaels, Antonio Davila, Scott Jackson, Ranjani Krishnan, Frank Moers, Daniel Oyon, Timothy Shields, Karen Sedatole, and Kristy Towry; workshop participants at Chapman University, Maastricht University, Michigan State University, Nanyang Technological University, the University of Arizona, and University of South Carolina, as well as participants at the American Accounting Association Annual Meeting, the Management Accounting Midyear Meeting, and mEmory Accounting Research Conference. The authors acknowledge financial support from the Goizueta Business School.

Aggregation Fallacy: An Explanation for Systematic Forecast Bias

Abstract:

In forecasting and capital budgeting, managers tend to underestimate costs. Prior research typically attributes this forecast bias to misaligned incentives or imperfect information. We introduce a novel explanation: *aggregation fallacy*. The key idea is that individuals provide forecasts of aggregate outcomes (e.g., total costs) by erroneously adding component forecasts (e.g., costs in divisions A and B). While such linear aggregation is appropriate for deterministic information (i.e., costs that have been incurred), it can result in systematic bias with probabilistic information (i.e., future costs) due to right skew in cost distributions. In support of our hypotheses, results of several experiments show that aggregation fallacy leads to systematic underestimation of costs. Additional analyses demonstrate that aggregation fallacy is not eliminated by learning and also holds with financial incentives for forecast accuracy. Overall, our study demonstrates that, even with aligned incentives and perfect information, systematic cost underestimation may arise due to aggregation fallacy.

JEL Classification: C44, C46, D24, D81, D91, G31, M40.

Keywords: Forecasting, capital budgeting, aggregation, fallacy, forecast bias.

I. INTRODUCTION

When preparing forecasts, managers tend to systematically underestimate costs, impairing planning and the efficient allocation of capital in firms (Adams 2024; Flyvberg, Holm, and Buhl 2002, 2003; Gray 2024; Jackson, Salzsieder, and Schaefer 2025; Klakegg and Lichtenberg 2016; Love et al. 2015; Merrow 2011). Prior research typically focuses on misaligned incentives and/or insufficient information to explain cost underestimation (Brüggen and Luft 2011, 2016; Haka 2006; Lovallo and Kahneman 2003). In this study, we show that, even with aligned incentives and perfect information, individuals systematically underestimate costs due to a fallacy in the aggregation of probabilistic information, that we coin *aggregation fallacy*.

To illustrate aggregation fallacy, consider a manager who must forecast the *total cost* of a project consisting of two cost components, A and B. When the costs of A and B are deterministic (i.e., they have already been incurred), aggregation is easy: add them up! Yet, when the costs of A and B are probabilistic (i.e., they will be incurred in the future), aggregation is more difficult. In particular, probabilistic features, such as the mode and median, that are often used to forecast individual cost components cannot simply be added up. That is, mode and median are not linear: mode(A) + mode(B) ≠ mode(A + B) and median(A) + median(B) ≠ median(A + B).

We predict that managers fail to account for these intricacies when deriving forecasts of total costs, as they tend to think in linear patterns (Larrick and Soll 2008; Stango and Zinman 2009). Specifically, we expect managers to fall prey to aggregation fallacy in that they derive aggregate forecasts by adding non-linear features of the component distributions. Importantly, aggregation fallacy leads to the systematic underestimation of costs, because costs tend to be right-skewed (in that mode < median < mean). Right skew in costs stems from a lower bound of zero without corresponding upper bound, producing long right tails. Empirical evidence across a range of industries supports the notion that costs are right-skewed (e.g., Amar 2025; Barber and Thompson 1998; Briggs and Gray 1998; Elliot and Payne 2005; Flyvberg et al. 2018; Karlsson, Wang, and Ziebarth 2024; Manju, Candel, and van Breukelen 2021).[1]

With right-skewed cost components, two key complexities arise that lead to aggregation fallacy and the systematic underestimation of total cost. First, the probabilistic features typically used for forecasting are no longer identical with right skew: mode < median < mean. Therefore, managers must decide whether they want to forecast the feature that (a) reflects the most-likely future outcome (mode), (b) minimizes the absolute forecast error (median), or (c) minimizes the squared forecast error (mean) (e.g., Barefield 1979). Prior research shows that individuals faced with this implicit decision tend to forecast the mode or median (Basu and Markov 2004; Chen, Kravet, and Ren 2025; Gu and Wu 2003), even given monetary incentives to forecast the mean (Petersen and Miller 1964).[2]

More strikingly, when managers aim to forecast the mode or median of *total cost*, a second complexity arises in that adding the modes [medians] of right-skewed cost components results in the systematic underestimation of the mode [median] of total cost. In particular, the central limit theorem dictates the following relations for right-skewed cost components: $\mathrm{mode(A)} + \mathrm{mode(B)} \leq \mathrm{mode(A + B)}$ and $\mathrm{median(A)} + \mathrm{median(B)} \leq \mathrm{median(A + B)}$. Thus, managers who fall prey to aggregation fallacy – i.e., deriving forecasts of total cost by adding the cost component modes [medians] – will underestimate not only the mean but also the mode [median] of total cost.

To test our theory, we also consider the hypothetical case of left-skewed cost components in that mode $>$ median $>$ mean. If cost components were left-skewed, the relations above would

[1] Right skew in cost distributions has been documented for the healthcare, construction, and aerospace industries. We also find strong evidence for right skew in costs with the universe of Compustat firms (see Section II for detail).
[2] Note that research in accounting expects managers to forecast means (Christensen 2010), because the mean is the only measure of central tendency that is always linear, and linear aggregation is prevalent in accounting.

reverse: $\text{mode(A)} + \text{mode(B)} \geq \text{mode(A + B)}$ and $\text{median(A)} + \text{median(B)} \geq \text{median(A + B)}$. Thus, managers who fall prey to aggregation fallacy would overestimate total costs. In sum, we predict that aggregation fallacy leads to (i) systematic underestimation when the components are right-skewed and (ii) systematic overestimation when the components are left-skewed.

To test our predictions, we first conduct two laboratory experiments (1a and 1b) in which we ask participants to provide their best forecast for the total cost of a project that consists of two cost components. Participants have perfect information about the cost components' distributions and no financial incentives to distort their forecasts. In each experiment, we manipulate the skew of the two cost components (both right-skewed versus both left-skewed).

The results of Experiment 1a support our theory. We find systematic underestimation when the two components are right-skewed and systematic overestimation when they are left-skewed. While we find some evidence for learning (i.e., reduction of bias) over time, improvements taper off, such that both forecast biases persist across all ten rounds of the experiment. In support of aggregation fallacy as the underlying mechanism, we find that a significant portion of forecasts is derived simply by adding the two component modes. Experiment 1b demonstrates that these forecast biases hold with a more experienced sample and provides more direct evidence of aggregation fallacy. After participants give their best forecast, we separately ask them to forecast the aggregate mode and aggregate median. The overwhelming majority – close to 90 percent – of participants derives these forecasts by adding the modes or medians of the components.

Results of three additional experiments (2a, 2b, and 2c) demonstrate that aggregation fallacy holds with simple, intuitive component distributions and explicit incentives for forecast accuracy. We conclude that the fallacy does not stem from a lack of motivation or the cognitive complexity of the component distributions being summed.

Our experiments also provide some support for an intervention to mitigate aggregation fallacy. Specifically, making salient to participants the midpoint of the aggregate distribution (i.e., information they could easily derive from the components) can limit aggregation fallacy by establishing a useful reference point to undo forecast bias. When the components are right-skewed, the aggregate midpoint is larger than the sum of the component modes (medians), such that participants may shift their forecasts upwards, thereby mitigating underestimation. In contrast, when the components are left-skewed, the aggregate midpoint is smaller than the sum of the component modes (medians), such that participants may shift their forecasts downwards, thereby mitigating overestimation.

Overall, we provide evidence for a fallacy in the aggregation of probabilistic information – i.e., *aggregation fallacy* – that can explain systematic underestimation of costs. Our results show that many individuals erroneously assume that probabilistic information can be aggregated in a manner similar to deterministic information, as they simply add up component modes or medians to derive aggregate forecasts. Doing so leads to biased forecasts, since neither mode nor median is linear. This misperception about the aggregation of probabilistic information offers a novel explanation for systematic forecast bias and suggests that aligning incentives and/or providing perfect information may not be sufficient to fully eliminate such bias. Our study answers recent calls for more accounting research on forecasting (Labro 2025) and highlights that organizations need to carefully consider different sources of forecast bias.[3]

In practice, the consequences of aggregation fallacy may be even more pronounced than in our stylized experiments. First, while we focus on the aggregation of only two cost components,

[3] Other examples of linear systems with consistently skewed component distributions include project time schedules (i.e., duration distributions tend to be right-skewed due to lower bounds at zero), the balance sheet (i.e., conservatism dictates the use of the least prudent of the most-likely outcomes), and asset pricing models (i.e., the distributions of stock returns tend to be right-skewed due to lower bounds at -100 percent).

large organizations disaggregate forecasts into hundreds or thousands of cost components, such that aggregation errors compound. Second, linear aggregation is embedded into the forecasting and capital budgeting processes of many organizations. For example, forecasts are often created using spreadsheets that simply add forecasts across various cost components without considering the possibility of non-linearities. Third, while participants in our study have perfect information about the cost component distributions, managers in practice often only obtain point forecasts. This lack of information about component distributions may make it more difficult to adjust for non-linearities in aggregating probabilistic information, and simply adding forecasts may feel even more natural when no other information is provided. In sum, we expect aggregation fallacy to be deeply ingrained in organizational forecasting processes; and our results suggest that a nontrivial share of managers may not be aware of the issue.

The remainder of this manuscript is organized as follows. Section II provides a theoretical underpinning and outlines our hypotheses. In Section III, we describe the experimental setting, design, and procedures. Sections IV, V, and VI present the results, and Section VII concludes.

II. BACKGROUND AND THEORY

Most accounting systems, including cost systems, are linear in that total system outcomes, like total costs, are calculated by adding all components in the system (e.g., all cost components). For example, total cost in the general ledger is the sum of all costs incurred by an organization; the total cost of a manufacturing division is derived by adding the costs of all products made in that division; and the total cost of a product is the sum of all costs assigned to that product. Such linear aggregation is appropriate when costs have been incurred. Yet, when forecasting costs, as in capital budgeting, the aggregation of individual cost forecasts into a forecast of total cost is more complex, because it involves probabilistic information.

In what follows, we outline these complexities and develop theory to suggest that managers do not fully account for them, resulting in systematic forecast bias, i.e., underestimation of costs, even with perfect information and absent any strategic motives.

Background

In forecasting and capital budgeting, costs are uncertain, reflecting probabilistic information. That is, any future cost can be considered a random variable following a probability distribution.[4] Managers are expected to provide cost forecasts that are representative of the underlying cost distributions, typically by estimating a measure of central tendency. The three most common measures of central tendency are (i) mode, (ii) median, and (iii) mean (e.g., Hogg et al. 2012).[5] For costs, the mode and median are usually smaller than the mean, because cost distributions tend to be right-skewed in that mode < median < mean.

At a theoretical level, prior research conjectures that costs are right-skewed because they have a natural left bound at zero – i.e., costs cannot be negative – but no corresponding right bound, thereby extending the right tail of the distribution (Berry and Otley 1975; Turvey 1971). Empirical evidence supports this line of reasoning. For example, evidence from the healthcare, construction, and aerospace industries shows right skew in cost distributions (e.g., Amar 2025; Book 2001; Briggs and Gray 1998; Elliot and Payne 2005; Flyvberg et al. 2018; Karlsson, Wang, and Ziebarth 2024).[6] To provide additional large-scale evidence, we also assess skew in quarterly Costs of Goods Sold (COGS) for the universe of firms in Compustat from 1962 to 2024 (34,558

---

[4] The probability distribution of a given cost is a mathematical description of the underlying uncertainty, specifying probabilities for the possible future costs that may be incurred.
[5] The mode represents the most-likely outcome of an underlying probability distribution; the median represents the outcome that splits the distribution into two halves of equal probability mass (i.e., the 50th percentile); and the mean represents the probability-weighted average of all possible outcomes (Hogg et al. 2012).
[6] Similarly, right skew in time distributions is well established (e.g., Mohan, Gopalakrishnan, Balasubramanian, and Chandrashekar 2007; Udoumoh and Ebong 2017). Because most costs are positively correlated with time (Lovallo and Kahneman 2003; Kaplan and Anderson 2008), right skew in time directly translates into right skew in costs.

firms and 5,089,724 firm-quarters). As expected, skew in COGS is significantly larger than zero (skew = 11.60, z = 9.76, $p < 0.001$, untabulated), reflecting right skew.[7]

This tendency toward right skew is critical, because it creates complexities that lead to the systematic underestimation of costs. First, with right skew, key distributional features typically used for forecasting – i.e., mode, median, and mean – differ, requiring managers to (implicitly) make a decision about the objective they want their forecasts to achieve. Forecasting the mode maximizes the likelihood of being exactly right. Forecasting the median minimizes the *absolute* forecast error, the absolute deviation of the forecast from the actual cost. This objective may be particularly intuitive, because, in practice, accuracy is often evaluated by how closely forecasts reflect actual costs. Finally, forecasting the mean minimizes the *squared* forecast error, ensuring a forecast of the probability-weighted average of all possible costs.[8]

When faced with this implicit decision, prior research suggests that managers often opt to forecast the mode or median. For example, analysts and managers make predictions as if they aim to forecast the median (Basu and Markov 2004; Chen, Kravet, and Ren 2025; Gu and Wu 2003). This tendency may reflect the notion that forecasting means (minimizing squared errors) is less intuitive than forecasting medians (minimized absolute errors) or modes (aiming to be exactly right).[9] Further, the mode and median may be considered more representative of skewed distributions than the mean, given the mean's sensitivity to outliers (e.g., Lock et al. 2021, 79).[10]

---

[7] Our analysis controls for firm size (via the natural log of total assets) while using firm-fixed effects to capture intra-firm variation in COGS across time and year-fixed effects to account for systemic changes over time.

[8] For symmetric, unimodal distributions like the normal distribution, mode, median, and mean are identical, such that the three objectives motivate the same forecast. Yet, when distributions have (right) skew, as is the case for costs, the three objectives motivate different forecasts.

[9] Peterson and Miller (1964) find that, even when individuals are given monetary incentives to forecast the mean of an underlying distribution, they tend to forecast the median. One potential explanation may be that it is difficult to internalize a *squared* penalty function, especially relative to an *absolute* penalty function.

[10] For example, when assessing household income, both mode and median are considered more meaningful than the mean – given the mean's sensitivity to large outliers. Similarly, when forecasting the weather, individuals may be more interested in the mode, the most-likely weather, than the mean (e.g., Engelberg, Manski, and Williams 2009).

The second complexity arising from right skew in costs – and the main focus of our study – is that, when managers intend to forecast the mode or median of total cost, the aggregation of individual cost forecasts into a forecast of total cost is not linear. To illustrate this non-linearity, consider a simple example in which two random variables – i.e., cost components A and B – are described by spinning two roulette-style wheels. The number each wheel lands on represents the realized cost of that component (A and B), and the sum of those two numbers represents the total cost (A + B). Before independently spinning each wheel, the actual costs are unknown, but they can be described by the probability distributions associated with each wheel. Figure 1 shows the distributions of the two wheels – i.e., the distributions of the two cost components – as well as the aggregate distribution – i.e., the distribution of total cost. Each wheel has 20 wedges, labeled with either a one (3 of 20 wedges: 15 %), a two (8 wedges: 40 %), a three (5 wedges: 25 %) or a four (4 wedges: 20 %). The distribution of each wheel is right-skewed in that mode and median (both 2) are smaller than the mean (2.5).

Comparing the two component distributions to the aggregate distribution shows that, while the mean is linear, mode and median are not. Specifically, the mean of the aggregate distribution (5) equals the sum of the component means (2.5). However, adding component modes [medians] (2 + 2) does not produce the mode [median] of the aggregation distribution (5). Thus, a manager who forecasts total cost by adding the component modes [medians] of A and B (2 + 2 = 4) will *underestimate* the aggregate mode [median] (5) as well as the aggregate mean (also 5).

The example above represents a generally applicable rule when the distributions of two cost components – here, A and B – are right-skewed, such that $\text{mode}(A) < \text{median}(A) < \text{mean}(A)$ and $\text{mode}(B) < \text{median}(B) < \text{mean}(B)$. In this case, the sum of the component modes [medians] falls short of the aggregate mode [median] and aggregate mean, as reflected in the following relations:

$$\text{mode(A)} + \text{mode(B)} \le \text{mode(A + B)} \quad \text{and} \quad \text{mode(A)} + \text{mode(B)} \le \text{mean(A + B)},$$

$$\text{median(A)} + \text{median(B)} \le \text{median(A + B)} \quad \text{and} \quad \text{median(A)} + \text{median(B)} \le \text{mean(A + B)}.$$

Because the cost components are right-skewed, simply adding the component modes [medians] leads to systematic underestimation of the aggregate mode [median] – and mean – of total cost. In other words, the aggregate distribution of total cost exhibits less right skew than the individual distributions of the cost components (as illustrated in Figure 1).

Note that the relations outlined above follow from the *Central Limit Theorem* and extend to the aggregation of any *n* cost components that are, on average, right-skewed, in that the sum of all *n* modes [medians] is smaller than the sum of all *n* means.[11] This tendency in skew is critical, since it requires managers to understand the non-linearities of probabilistic features, such as the mode and median, when aggregating forecasts of individual costs into a forecast of total cost.

We develop theory to suggest that managers fail to understand these non-linearities, leading to systematic forecast bias. Note that our contribution is not to outline *analytically* complexities in the aggregation of skewed component distributions.[12] Rather, our contribution is to provide

[11] The *Central Limit Theorem* implies that the sum of any n i.i.d. random variables, X_i, where $i \in \{1, \ldots, n\}$, converges to an approximately normal distribution (e.g., Hogg et al. 2012). Importantly, the theorem holds irrespective of the variables' distribution and skew. Assuming that X_i is right-skewed with $\text{mode}(X_i) < \text{median}(X_i) < \text{mean}(X_i)$, we can deduce that the sum of the individual modes is smaller than the sum of the individual medians, which is smaller than the sum of the individual means: $\sum_i(\text{mode}(X_i)) < \sum_i(\text{median}(X_i)) < \sum_i(\text{mean}(X_i))$. Since we know that (i) the mean is linear, $\sum_i(\text{mean}(X_i)) = \text{mean}(\sum_i(X_i))$, and (ii) both aggregate mode and median – of the sum – converge toward the aggregate mean, we can derive the following relations:

$$\sum_i(\text{mode}(X_i)) \le \text{mode}(\textstyle\sum_i(X_i)) \text{ and } \sum_i(\text{mode}(X_i)) \le \text{mean}(\textstyle\sum_i(X_i)), \text{ and}$$
$$\sum_i(\text{median}(X_i)) \le \text{median}(\textstyle\sum_i(X_i)) \text{ and } \sum_i(\text{median}(X_i)) \le \text{mean}(\textstyle\sum_i(X_i)).$$

For ease of illustration, we have assumed i.i.d. random variables. Yet, probability mass still converges toward the sum of means – and, hence, the relations outlined above still hold – when the variables are not independent of each other and/or not identically distributed. For example, when independence is not given, pace of convergence toward the sum of means depends on the correlations among the variables. Positive correlations slow down convergence; negative correlations speed up convergence. Unless all variables are perfectly, positively correlated, convergence of probability mass toward the sum of means occurs. In the case of perfect, positive correlations among all variables, individual modes [medians] add up to the aggregate mode [median], and there is effectively one random variable. When the random variables are not identically distributed, probability mass still converges toward the sum of means and, in many cases, convergence is toward an approximately normal distribution (Andrews 1991).

[12] Non-linearity of mode and median is well established and follows from the central limit theorem. In accounting, Otley (1985) algebraically illustrates that, when distributions of performance targets are skewed, the distribution of

theory and empirical evidence that managers do not understand these complexities (i.e., non-linearities), and that this lack of understanding leads to the systematic underestimation of costs, even with perfect information about the components and absent strategic motives.[13]

Hypothesis Development

Our main argument is that managers fall prey to *aggregation fallacy*, the mistaken belief that adding a (non-linear) probabilistic feature, such as the mode or median, across component distributions will produce the same feature of the aggregate distribution. For example, a manager who falls prey to aggregation fallacy will assume that the mode [median] of a project's total cost can be derived by adding the modes [medians] of the cost distributions of all project components.

Our theory builds on the notion that individuals tend to think in linear patterns and struggle to understand non-linearities (e.g., de Langhe, Puntoni, and Larrick 2017; Larrick and Soll 2008; Stango and Zinman 2009). Such tendencies are likely reinforced when forecasting costs, because managers are used to the fact that costs *can* be aggregated linearly, once they have been incurred. Thus, we expect managers to assume that the mathematical rules they apply to deterministic cost information – i.e., adding costs that have been incurred – can also be applied to probabilistic cost information – i.e., adding forecasts of costs that will be incurred.

As outlined above, this assumption is inaccurate for forecasts of modes and medians, but it

---

their sum will be less skewed. Moreover, Schwartz, Spires, Wallin, and Young (2012) examine how the non-linearity of the standard deviation influences strategic interactions in a budgeting setting.

[13] Prior research provides survey evidence for conceptual misconceptions about the definition and properties of the *Central Limit Theorem* among social science researchers (e.g., Zhang et al. 2023). These conceptual misconceptions relate to technical aspects of the theorem – unrelated to the (non-)linearity of common measures of central tendency. In contrast, our study focuses on an intuitive understanding of the linearity of the mean and the non-linearity of the mode and median, which follows from the *Central Limit Theorem* (i.e., the theorem implies that the aggregate mode and median converge toward the aggregate mean). Perfect understanding of the theorem is not a necessary condition to understand the non-linearity of mode and median, just as a perfect understanding of *Bayes' Theorem* is not needed for individuals to behave in line with *Bayesian* updating. Contrary to research that studies the heuristics individuals may apply to determine or update the probability of a given event (e.g., Juslin, Lindskog, and Mayerhofer 2015), our study focuses on the aggregation of probabilistic information across multiple events, absent any updating.

is accurate for forecasts of means. Thus, managers can avoid systematic forecast bias by adding component means, because doing so necessarily produces the aggregate mean due to the linearity of the mean. Yet, when distributions are skewed, individuals often view the mode and/or median as more representative than the mean (e.g., Basu and Markov 2004; Chen, Kravet, and Ren 2025; Petersen and Miller 1964; Gu and Wu 2003). Therefore, we argue that, even if component means are available, managers often aim to forecast the mode or median of total cost, thereby making themselves prone to aggregation fallacy. As such, we expect them to derive forecasts of the mode [median] of total cost by adding the component modes [medians]. Since costs are right-skewed, such linear aggregation of component modes [medians] leads to the systematic underestimation of the mode [median] – and mean – of total cost (as outlined in the analytical relations above).

In sum, we expect that managers often aim to forecast the mode or median of total cost and, consequently, fall prey to aggregation fallacy by adding component modes or medians to derive their forecasts. Thus, we hypothesize that, when component distributions are right-skewed, as is the case for costs, individuals will understate forecasts of aggregate system outcomes. Formally, we state the following hypothesis:

H1a: When component distributions are right-skewed, forecasts of aggregate system outcomes will be understated.

While costs are typically right-skewed in practice, we also consider a hypothetical case with left-skewed components. Doing so enables a direct process test, because our theory suggests that the direction of skew alters the sign of forecast bias. That is, with left skewed components, we expect overestimation due to aggregation fallacy. Our reasoning is identical to that above, with the key difference that, when the components are left-skewed, in that mode > median > mean, the analytical relations reverse:

$$\text{mode}(A) + \text{mode}(B) \geq \text{mode}(A + B) \quad \text{and} \quad \text{mode}(A) + \text{mode}(B) \geq \text{mean}(A + B),$$

$$\text{median}(A) + \text{median}(B) \geq \text{median}(A + B) \quad \text{and} \quad \text{median}(A) + \text{median}(B) \geq \text{mean}(A + B).$$

Thus, we state the following hypothesis:

H1b: When component distributions are left-skewed, forecasts of aggregate system outcomes will be overstated.

Note that we implement a conservative test of our theory by examining aggregation fallacy when individuals have perfect information about the cost component distributions. We use this stylized setting to isolate misperceptions about the aggregation of probabilistic information from misperceptions about the information itself. In practice, avoiding aggregation fallacy is likely more cognitively challenging, because perfect information about the component distributions is rarely available.

Next, we consider an intervention to mitigate aggregation fallacy. Given the complexities in the aggregation of probabilistic component information, we argue that managers might benefit from focusing on simple features of the *aggregate* distribution. Specifically, directing managers' attention to the aggregate midpoint (and range) of total cost – information that can be obtained with relative ease – might mitigate aggregation fallacy by providing a useful reference point.[14]

When the component distributions are right-skewed, the midpoint of the aggregate range is (weakly) larger than both the sum of the component modes [medians] and the mode [median] of the sum. Thus, managers who add components modes or medians observe a discrepancy from the aggregate midpoint. We argue that this discrepancy will lead managers to question their strategy and revise their forecasts upward toward the aggregate midpoint, thereby reducing discrepancy

[14] In contrast to component means, we expect that the minimum and maximum costs of individual components can be obtained with relative ease from domain experts, allowing organizations to derive the minimum and maximum of total cost and, thus, make salient to managers the aggregate midpoint (and range) of total cost. Note that, except for the extreme case of a perfect negative correlation among the components, minimum and maximum are linear.

and counteracting underestimation. We formally state the following hypothesis:

H2a: When component distributions are right-skewed, increased salience of the midpoint and range of the possible aggregate outcomes will mitigate understatements in forecasts.

In contrast, when the component distributions are left-skewed, the midpoint of the aggregate range is (weakly) smaller than both the sum of the component modes [medians] and the mode [median] of the sum. To reduce this discrepancy, we expect managers to revise their forecasts downward, thereby counteracting overestimation. Formally, we state the following hypothesis:

H2b: When component distributions are left-skewed, increased salience of the midpoint and range of the possible aggregate outcomes will mitigate overstatements in forecasts.

III. METHODS OVERVIEW

To test our hypotheses, we conduct multiple experiments in which participants forecast an aggregate outcome that is the sum of two random components, A and B.[15] Participants are given the distributions of the components and are informed that they are independent from each other. They can use this information to infer the distribution of the sum (i.e., the distribution of A + B). We give participants the component distributions to isolate misperceptions about the aggregation of probabilistic component information from misperceptions about the components themselves.

In each experiment, we manipulate the skew of the component distributions, such that both components have either right skew, reflecting typical cost distributions in practice, or left skew. If individuals fall prey to aggregation fallacy, they will systematically underestimate with right skew and overestimate with left skew. Our main dependent variable of interest is forecast bias. While practitioners may consider forecasting the median (i.e., minimizing absolute deviations) a meaningful objective (Basu and Markov 2004; Chen, Kravet, and Ren 2025; Gu and Wu 2003), linear accounting systems imply managers should forecast the mean (i.e., minimizing squared

[15] All experiments received required IRB approval. Note that the instructions refer to *estimates* rather than *forecasts*.

deviations) (e.g.. Christensen 2010). Thus, we use both benchmarks and calculate forecast bias as the forecast of the sum minus the mean or median of the distribution of the sum.[16]

In what follows, we distinguish two sets of experiments. Experiments 1a and 1b investigate aggregation fallacy in a context-rich, capital budgeting setting. Experiments 2a, 2b, and 2c study aggregation fallacy in a more abstract setting with monetary incentives for forecast accuracy.

IV. EXPERIMENT 1A

In Experiment 1a, participants assume the role of a financial analyst who must forecast the total cost of a capital budgeting project, which is the sum of the costs of two project components. Participants are asked to provide, in each of ten rounds, their *best* forecast of total project cost.[17] While there are no explicit monetary incentives for forecast accuracy, the work scenario implies that accuracy is important, and participants have no incentives to distort their forecasts.

We give participants perfect information about the two cost components by providing them with textual and visual depictions of the component distributions. Each cost component follows a triangular distribution. We use triangular distributions because they are common in cost analysis (e.g., Johnson 1997; Banks et al. 2004; Mun 2006) and simple to understand, as they are defined by three parameters: minimum, mode, and maximum.[18] Participants observe the three parameters for each component, allowing them to derive its mean and median.[19] Note that in Experiment 1b,

---

[16] As will be described, aggregate mode and median are similar (identical) in Experiments 1a and b (2a, b, and c), such that inferences derived from using the median as benchmark also hold when using the mode as benchmark. Note that forecasting the mode maximizes the likelihood of being exactly right, which approaches zero in practice, where distributions are continuous, and thus has little meaning relative to minimizing absolute or squared deviations.
[17] Prior research notes that, in practice, forecasters are usually asked to provide a "best prediction" rather than a forecast of the mode, median, or mean (Engelberg, Manski, and Williams 2009, 30; Jørgensen 2014).
[18] In practice, a key benefit of triangular distributions is that minimum, mode, and maximum can be elicited from experts with relative ease compared to other features – e.g., mean, variance, or kurtosis – needed to define more complex distributions. In the experiment, a key benefit of triangular distributions is that we can manipulate skew (right versus left) by shifting the mode below versus above the midpoint, holding constant minimum and maximum.
[19] The mean of a triangular distribution is the average of its minimum, mode, and maximum. While somewhat more complex, the median of a triangular distribution can also be calculated based on its minimum, mode, and maximum.

we also explicitly tell participants the component means and medians. To ensure understanding, the instructions include a primer explaining triangular distributions and concepts, such as the mode, median, and mean. The instructions also include eleven understanding checks, of which six pertain to participants' understanding of probabilistic information (from the primer).

Across all ten (project) rounds, the distributions of the two cost components – and, thus, the aggregate distribution of total project cost – are identical. To assess whether participants learn to avoid aggregation fallacy with experience, they receive feedback after each forecast. That is, at the end of each round, participants observe their forecast of total cost and the actual total cost, as determined by random draws from the two cost component distributions. Using repeated choices with identical distributions (across ten rounds) provides a stark test of learning. In the workplace, learning is more complex, because employees rarely forecast the same (project) cost repeatedly, and economic conditions evolve with time.

Within this setup, we use a 2 x 2 between-participants design that applies to all ten rounds. First, we manipulate the skew of the component distributions (both right skew versus left skew) by varying the modes – while holding the minima and maxima constant. With right skew, we shift the modes below the midpoints of the distributions: mode < median < mean < midpoint. With left skew, we shift the modes above the midpoints: mode > median > mean > midpoint. Irrespective of skew, the two components only differ in that we add a fixed amount of $5,000 to B relative to A. Figure 2 shows the distributions of the two components and their sum with both right skew (Panel A) and left skew (Panel B). Participants do not observe the distribution of the sum. A naïve decision-maker might assume that the sum (i.e., the total project cost) also follows a triangular distribution. Yet, irrespective of skew, the distribution of the sum is close to normal, as probability mass and, thus, aggregate mode and median approach the aggregate mean.

Notably, with right [left] skew, the sum of the component modes, 96,000 = 45,500 + 50,500 [144,000 = 69,500 + 74,500], is smaller [larger] than the aggregate mode of 110,237 [130,091]. Similarly, the sum of the component medians, 109,000 = 52,000 + 57,000 [131,000 = 63,000 + 68,000], falls short of [exceeds] the aggregate median of 111,130 [128,870]. In contrast, the mean is linear, with the sum of component means, 112,000 = 53,500 + 58,500 [128,000 = 61,500 + 66,500] equaling the aggregate mean, 112,000 [128,000], irrespective of skew.

Second, we manipulate the salience (low versus high salience) of the midpoint and range of the distribution of the sum. In the *baseline* condition (low salience), participants do not receive explicit information about the distribution of the sum. In the *treatment* condition (high salience), we explicitly inform participants about the aggregate midpoint and range. Note that participants can always infer this information from the component minima and maxima.

Participants

In total, 504 BBA students at a U.S. business school took part in Experiment 1a in exchange for course credit. Of the 504 participants, 305 (60.5%) are male, and the average age is 19.81 years. It took participants an average [median] time of 11.93 [8.74] minutes to complete the task.

Results

We restrict our main analyses to the 346 participants (68.7%) who answered all attention checks correctly and, for ease of presentation, focus primarily on forecast bias using the aggregate mean as the benchmark. Yet, we also report results for the full sample and for forecast bias relative to the median. In both cases, the key inferences remain largely unchanged.

Table 1 and Figure 3 (Panel A) depict forecast bias, relative to the mean, averaged across all ten rounds, separately for each condition.[20] We find negative forecast bias (underestimation) with

[20] From the results reported in Table 1, the interested reader can easily derive the corresponding forecast bias relative to the median, since the aggregate median falls short of [exceeds] the aggregate median by 870 with right [left] skew

right skew and positive forecast bias (overestimation) with left skew, supporting our main theory.

To test our hypotheses, we employ the following regression model and cluster standard errors by participant, *i*, accounting for interdependence of observations across the ten project rounds, *r*:

$$Forecast\ Bias_{ir} = \beta_0 + \beta_1\ Left\ Skew_i + \beta_2\ High\ Salience_i + \beta_3\ Left\ Skew_i \times High\ Salience_i + \varepsilon_{ir},$$ where:

- $Forecast\ Bias_{ir}$ is the deviation of participant i's forecast in project round *r* from the mean (or median) of the aggregate distribution of total project cost,
- $Left\ Skew_i$ is one (zero) for projects with components that have left (right) skew, and
- $High\ Salience_i$ is one (zero) for high (low) salience of the aggregate midpoint and range.

Table 2, Panel A shows the results. Column 1 [2] depicts the results for forecast bias relative to the mean [median].[21] Focusing first on the baseline condition (low salience), we find that skew significantly affects forecast bias (column 1: β_1 = 9,161, p < 0.001).[22] In line with our prediction, participants underestimate total costs when components are right-skewed (H1a: *Forecast Bias* = β_0 = -4,213, p < 0.001) and overestimate total costs when components are left-skewed (H1b: *Forecast Bia*s = β_0 + β_1 = 4,948, p < 0.001). This pattern is consistent with aggregation fallacy, linking the sign of forecast bias to the direction of skew.

Focusing on the treatment condition (high salience), we still find that skew affects forecast bias (β_1 + β_3 = 5,787, p < 0.001), but its impact is significantly less pronounced than in the baseline condition (β_3 = -3,373, p = 0.087). As such, our results provide evidence that increasing salience of the aggregate midpoint and range reduces forecast bias. Yet, this effect of salience only materializes with left skew. That is, increasing salience of the aggregate midpoint and range does not significantly attenuate underestimation with right skew (H2a: p = 0.942), but it does

---

(see Figure 2). Thus, with right [left] skew, adding 870 to [subtracting 870 from] the reported forecast bias produces the forecast bias relative to the median. Importantly, the standard deviations remain the same.

[21] Note that the coefficients in columns 1 and 2 differ only for *Left Skew* (by 1,740 = 2 * 870) and *Constant* (by 870), where 870 is the difference between aggregate mean and median. Similarly, forecast bias in each condition differs by this difference of 870 between columns 1 and 2. Importantly, standard errors are identical across columns 1 and 2.

[22] All reported p-values are two-tailed.

significantly attenuate overestimation with left skew (H2b: $p = 0.013$). Thus, the intervention mitigates the effect of skew – a critical consequence of aggregation fallacy – but is not effective at reducing underestimation with right skew, the typical case for costs. We revisit this result in Experiments 2a, 2b, and 2c.[23]

Figure 4 illustrates that, irrespective of skew, forecast bias persists across all ten rounds.[24] While forecast bias is significantly reduced from round 1 to round 2, we do not find significant changes between rounds 2 and 10.[25] Further, systematic underestimation [overestimation] with right [left] skew remains significant on average across rounds 2 to 10.[26] We conclude that, even in the extreme scenario of repeated exposure to ten identical projects with the same distributions, feedback and learning alone are insufficient to eliminate forecast bias due to aggregation fallacy.

Closer examination of participants' forecasts provides more direct support that they fall prey to aggregation fallacy: the mistaken belief that adding the mode [median] of each component produces the mode [median] of the sum of components. Specifically, we examine how often participants' forecasts of total cost equal the sum of the component modes

Figure 5 illustrates that approximately 20 percent of forecasts – across all ten rounds – equal the sum of the component modes (i.e., $96,000 with left skew and $144,000 with right skew).

---

[23] With the full sample, the first hypothesis is supported, as forecast bias significantly differs across the two skews (see Table 2, column 3): significant underestimation with right skew and overestimation – albeit not significant at conventional levels – with left skew. Notably, collapsing across the high and low salience conditions, overestimation with left skew becomes marginally significant with the full sample ($p = 0.098$, untabulated). The second hypothesis – i.e., salience of aggregate midpoint and range reduces forecast bias – is not supported with the full sample.

[24] For ease of illustration, we present forecast bias pooled across the low and high salience conditions. We confirm the results (without pooling) and find significant forecast bias for each condition, except for the condition with left skew and high salience, where the salience treatment is effective at reducing overestimation.

[25] Forecast bias is attenuated from round 1 to round 2 by 7,046 with right skew ($t = 4.13$, $p < 0.001$, $df = 345$) and by 4,535 with left skew ($t = -2.30$, $p = 0.022$, $df = 345$, both untabulated). Yet, there is no significant attenuation in bias from round 2 to round i, for any i, where $3 \leq i \leq 10$ (right-skew: smallest $p = 0.328$; left-skew: smallest $p = 0.168$); similarly, the average change from round 2 to round 10 is not significant with right skew ($p = 0.915$, $df = 345$) or with left skew ($p = 0.407$, $df = 345$, all un-tabulated).

[26] Across rounds 2 through 10, average forecast bias is -3,572 with right skew ($t = -4.68$, $p < 0.001$, $df = 345$) and 2,731 with left skew ($t = 3.69$, $p < 0.001$, $df = 345$, both un-tabulated). In round 10, forecast bias is -2,439 with right skew ($t = -1.73$, $p = 0.084$, $df = 345$) and 4,182 with left skew ($t = 3.74$, $p < 0.000$, $df = 345$, un-tabulated).

The proportion of forecasts that simply add component modes is significantly greater than zero in each of the four conditions (all $p < 0.001$, untabulated). Adding component modes is also tied to systematic forecast bias. That is, compared to other forecasts, forecasts that simply add modes exhibit greater underestimation with right skew ($\Delta = 7{,}279$, $t = 8.22$, $p < 0.001$, $df = 345$) and greater overestimation with left-skew ($\Delta = 9{,}988$, $t = 11.00$, $p < 0.001$, $df = 345$).

Interestingly, making the aggregate midpoint and range salient leads to a (marginal) increase in the proportion of forecasts that add up component midpoints to derive a forecast of total cost (right skew: $\Delta = 3.90\%$, $t = 1.51$, $p = 0.131$, $df = 345$; left skew: $\Delta = 5.76\%$, $t = 2.35$, $p = 0.019$, $df = 345$). This finding supports the notion that increasing the salience of the aggregate midpoint shifts participants' focus toward the center of the distribution.

In sum, Experiment 1a shows that many individuals fail to grasp the complexities involved in aggregating probabilistic information in the forecasting process. Aggregation fallacy leads to systematic underestimation when components are right-skewed, adding to our understanding of cost understatements (H1a), and to systematic overestimation when components are left-skewed, linking bias to skew, consistent with our theory (H1b). We next present Experiment 1b, which aims to (i) replicate these results with a more experienced sample – i.e., MBA students – and to (ii) more directly capture the strategies individuals use when providing aggregate forecasts.

V. EXPERIMENT 1B

Experiment 1b implements the same capital budgeting context as Experiment 1a but makes several changes to the experimental setup and design. First, in addition to the minima, maxima, and modes of the component distributions, we provide participants with the component means and, in some conditions, with the component medians. As such, we can rule out that participants incorrectly infer component means and/or medians.

Second, we employ a 2 x 2 mixed-design. We manipulate – within-participants – the skew of the component distributions (right skew versus left-skew) and – between-participants – whether or not the instructions also include component medians (with medians versus without medians). Participants forecast total costs first for the project with right-skewed components and then for the project with left-skewed components (without feedback). Because repeated forecasting did not eliminate forecast bias in Experiment 1a, we use a simpler design that requires participants to provide only one best forecast of total cost for each skew. Moreover, rather than increasing the salience of the aggregate midpoint and range, we manipulate whether participants also receive explicit information about the component medians. In the baseline condition (with medians), we explicitly tell participants the component medians. In the treatment condition (without medians), participants are not told the component medians, but they can always infer them.

Third, after providing their best forecast of total cost for each project, we ask participants to provide forecasts of the aggregate mode, median (only in the baseline condition), and mean. Those additional forecasts enable a direct test of participants' beliefs about the linearity of the three measures of central tendency.

Participants

In total, 131 MBA students at a U.S. business school participated in Experiment 1b in exchange for course credit. Of the 131 participants, 80 (61.1%) are male and the average age is 30.12 years. It took participants an average [median] time of 11.07 [10.46] minutes to complete the task. We again restrict our main analyses to the 93 (71.0%) participants who answered all eleven attention checks correctly. Table 2, Panel B also reports the results for the full sample.

Results

Figure 3 and Table 1 (Panel B) illustrate that forecast bias persists in Experiment 1b despite providing participants with information about the component means (and medians). Table 2, Panel B reports the results of regressing forecast bias on the experimental variables. We employ the same regression model as in Experiment 1a but replace the variable capturing the treatment effect with the variable *No Median* (in lieu of *High Salience*). Again, we cluster standard errors by participant, allowing us to account for interdependence of observations across the two skews.

In the baseline condition with medians, participants underestimate total cost with right skew (H1a: *Forecast Bias* = β_0 = -5,155, p < 0.001) and overestimate total cost with left skew (H1b: *Forecast Bias* = $\beta_o + \beta_1$ = 6,286, p < 0.001). We do not find that withholding the medians significantly mitigates forecast bias (β_3 = 1,236, p = 0.738). As such, in the treatment condition (without medians), we continue to find underestimation with right skew (*Forecast Bias* = $\beta_o + \beta_2$ = -6,657, p = 0.002) and overestimation with left skew (*Forecast Bias* = $\beta_o + \beta_1 + \beta_2 + \beta_3$ = 6,020, p = 0.038). Thus, forecast bias does not seem to stem from an inability to infer component means (or medians) but rather from a misperception about the aggregation of probabilistic information.

Figure 5, Panel B provides direct evidence for aggregation fallacy by showing that many individuals appear to be adding component modes or medians. Specifically, irrespective of skew, we find that more than 50 percent of participants in the baseline condition (with medians) and more than 30 percent in the treatment condition (without medians) provide forecasts that equal the sum of component modes or medians (all p < 0.001 relative to 0 percent, untabulated).

Figure 6 shows the different strategies participants utilize when asked specifically for their forecasts of the aggregate mode, (median,) and mean. The results provide striking evidence for aggregation fallacy. The vast majority of participants – between 79 and 95 percent – derive forecasts of the aggregate mode [median] by adding component modes [medians], suggesting they erroneously assume linearity. Similarly, we find that more than 80 percent of participants add the component means to derive forecasts of the aggregate mean. As such, our results suggest that emphasizing means rather than modes or medians may mitigate forecast bias. However, in practice, component means may be difficult to obtain, especially when component distributions are skewed.[27] Thus, emphasizing means may not be a viable option.

Overall, Experiments 1a and 1b show robust evidence for forecast bias stemming from aggregation fallacy. Despite perfect information about the component distributions, many participants forecast total costs by applying an intuitive but incorrect heuristic – i.e., adding the modes or medians of component costs – leading to systematically biased forecasts.

VI. EXPERIMENTS 2A, 2B, AND 2C

In Experiments 1a and 1b, the two components follow continuous (triangular) distributions. It is possible that participants find it difficult to understand continuous distributions, potentially contributing to aggregation fallacy. Moreover, the exact shape of distributions is rarely specified in practice. Rather, managers likely receive information that allows them to develop an intuitive understanding of the underlying component distributions. To this end, we design Experiments 2a, 2b, and 2c to provide participants with an intuitive understanding of the component distributions while retaining key elements of the budgeting context. Participants still forecast an aggregate outcome that is the sum of two components. Yet, we use a more intuitive context and simpler component distributions. That is, we operationalize the two components as roulette-style wheels, each with six wedges that are equally likely to occur and labeled with a number between 1 and 5, reflecting discrete distributions. The two wheels have identical skew and are independently spun, each determining a separate number. The objective is to forecast the sum of the two numbers.

---

[27] When historical data is available, sampling distorts estimates of component means toward modes and medians in insufficiently large samples. Similarly, when historical data is not available, and estimates of component means must be obtained from domain experts, cognitive biases will likely distort them toward modes and medians.

The graphic below illustrates the game when both wheels have right skew versus left skew. With right skew, on each wheel, the number 2 appears on two wedges, while 1, 3, 4, and 5 appear on only one wedge. In contrast, with left skew, on each wheel, 4 appears on two wedges, while 1, 2, 3, and 5 appear on only one wedge. As such, participants receive perfect information about the component distributions in a format that is more intuitive.

Figure 7 illustrates, for each skew, the distributions of the two components and their sum. With right skew, the two component modes [medians] are 2 [2.5], whereas with left skew, the two component modes [medians] are 4 [3.5]. However, irrespective of skew, the aggregate mode [median] is 6, illustrating the non-linearity of the mode [median]. Moreover, the aggregate mean is best approximated by 6, both with right skew (aggregate mean = 5.67) and with left skew (aggregate mean = 6.33). Therefore, regardless of (i) whether participants aim to forecast the aggregate mode, median, or mean and regardless of (ii) skew, it is optimal to forecast a sum of 6.

To ensure aggregation fallacy does not stem from a lack of cognitive effort, we also provide participants with monetary incentives for forecast accuracy. Specifically, we incentivize them to forecast (a) the aggregate mode in Experiment 2a, (b) the aggregate median in Experiment 2b, and (c) the aggregate mean in Experiment 2c.[28] In each experiment, forecasting a sum of 6 is

[28] In Experiment 2a, we incentivize forecasting the aggregate mode by giving an all-or-nothing bonus if the forecast equals the actual sum (determined by the two numbers independently spun on the two wheels). In Experiment 2b, we incentivize forecasting the aggregate median by giving a bonus that depends on the *absolute* deviation of the forecast from the actual sum: $Bonus = \$A - \$B \cdot |Forecast - Actual|$. Finally, in Experiment 2c, we incentivize forecasting the aggregate mean by giving a bonus that depends on the *squared* deviation of the forecast from the actual sum: $Bonus = \$C - \$D \cdot (Forecast - Actual)^2$. The exact parameters are presented in Appendix A.

optimal, because it maximizes the expected bonus (as noted above). Thus, we calculate forecast bias as participants' forecast minus the optimal forecast of 6.

In all three experiments, we use a 2 x 2 between-participants design. First, we manipulate the skew of the component distributions (both right skew versus left skew) by varying whether the number 2 or 4 is on each wheel twice, as stated above. Second, we manipulate the salience (low versus high) of the midpoint – and range – of the aggregate distribution. In the *low salience* condition, participants do not receive explicit information about the aggregate midpoint. In the *high salience* condition, we explicitly provide them with the aggregate midpoint (and range).[29]

Before participants provide their forecast, they can test the game by clicking a button that spins the wheels and determines a sum based on the two numbers randomly spun. Participants can test the game as often as they wish. Providing participants with this option (i) ensures their understanding of the game and (ii) allows them to develop an intuition about the distribution of the sum. After testing the game, participants submit their forecast for the actual game. They then initiate the "game spin" to determine their compensation based on the actual sum spun.

Participants

In total, we recruited 1,463 participants: 876 for Experiment 2a with mode incentives, 401 for Experiment 2b with median incentives, and 186 for Experiment 2c with mean incentives.[30] Of the 1,463 participants, 733 (50.1%) are male, and the average age is 25.18 years. It took them a median time of 10.60 minutes to complete the task. Again, we restrict the main analyses to the 1,112 participants (76.0%) who answered all attention checks correctly.

[29] We also include a condition in which we provide explicit information about the aggregate range but not about the aggregate midpoint. Forecast bias in this condition does not significantly differ from the *baseline* condition which does not provide any aggregate information. Thus, for ease of presentation, we pool the baseline condition and the condition that only provides the aggregate range (but not the aggregate midpoint) into one *low salience* condition.
[30] The data was collected with different populations, including BBA students (n = 807), MBA students (n = 403), and workers on Mechanical Turk (n = 253). See Appendix A for full details. The results do not significantly differ across those populations. Thus, we report the results pooled across the three populations for ease of presentation.

Results

Figure 8 and Table 3 show forecast bias, separately for each condition in each experiment. In all three experiments, we find negative forecast bias (underestimation) with right skew and positive forecast bias (overestimation) with left skew. Thus, our results provide strong support for an ingrained misperception about the aggregation of probabilistic information.

Table 4 shows the regression results, separately for each experiment (columns 1 to 3) and pooled across experiments (column 4). When salience of the aggregate midpoint is low, skew alters forecast bias in all three experiments (all $\beta_1 > 1.50$, $p < 0.001$). Specifically, participants underestimate the sum with right skew (all $\beta_0 < -0.65$ and all $p < 0.001$), and they overestimate the sum with left skew (all $\beta_0 + \beta_1 > 0.80$ and all $p < 0.001$). Thus, aggregation fallacy persists even when participants can develop an intuitive understanding of the distributions and are given monetary incentives to provide accurate forecasts, irrespective of the type of incentive.

We also find that salience of the aggregate midpoint mitigates forecast bias in Experiments 2a (column 1: $\beta_3 = -0.47$, $p = 0.020$) and 2c (column 3: $\beta_3 = -1.09$, $p = 0.003$). More specifically, in Experiment 2a, salience of the aggregate midpoint reduces underestimation with right skew ($\beta_2 = 0.38$, $p = 0.020$) but not overestimation with left skew ($\beta_2 + \beta_3 = -0.10$, $p = 0.433$). Only in Experiment 2c does making the midpoint salient mitigate both underestimation with right skew ($\beta_2 = 0.43$, $p = 0.036$) and overestimation with left skew ($\beta_2 + \beta_3 = -0.66$, $p = 0.026$). As such, support for the effectiveness of the intervention is mixed: making the aggregate midpoint salient *can* mitigate aggregation fallacy, as 2 [1] out of 3 effects are significant with right [left] skew.

Finally, Figure 9 illustrates that, when salience of the aggregate midpoint is low, 45 to 70 percent of participants provide forecasts that equal the sum of component modes or component medians. When salience of the aggregate midpoint is high, the percentage of participants falling prey to aggregation fallacy is only 24 to 62 percent. This drop is largely driven by fewer participants forecasting the sum of component modes.

Overall, aggregation fallacy appears to be a robust phenomenon even when individuals have perfect information and well-aligned incentives. Aggregation fallacy has important implications for managerial decision-making as it leads to systematic cost underestimation, given right skew in costs. Our results also suggest that interventions designed to improve forecast accuracy must go beyond providing better (component) information and/or aligning incentives.

VII. CONCLUSIONS

In this study, we introduce a novel source of forecast bias related to the aggregation of probabilistic information. In contrast to aggregating deterministic outcomes (historical costs), aggregating probabilistic information (future cost) is complex because common probabilistic features – such as mode and median – are not linear (e.g., mode[A+B] ≠ mode[A] + mode[B]). We develop theory to suggest many individuals fail to account for these complexities. Instead, they fall prey to *aggregation fallacy* – the erroneous belief that probabilistic information can be linearly aggregated like deterministic information. Importantly, aggregation fallacy results in predictable forecast bias because costs tend to be right skewed. With right-skew, the sum of the component modes [medians] falls short of the aggregate mode [median]. Thus, managers who forecast total cost by adding the modes [medians] of all cost components will systematically underestimate the mode [median] of the total cost. Results from a series of experiments support the theory, with participants systematically underestimating total costs when components are right-skewed. Pinpointing aggregation fallacy as the mechanism, additional analyses show that many mistakenly assume the mode and/or median can be linearly aggregated.

Overall, these findings show that forecast bias is not only driven by incentive misalignment or lack of information. Even with perfect information and incentives for accuracy, our study demonstrates how individuals produce biased forecasts due to the erroneous aggregation of probabilistic information. This insight suggests solely addressing incentive and/or informational issues may be insufficient, because aggregation fallacy can independently lead to forecast bias. Given the prevalence of linear aggregation – across different types of costs, products, and organizational units – the cumulative effect of aggregation fallacy can be substantial.

Our study makes several contributions to theory and practice. At a theoretical level, our findings call for closer scrutiny of the "master-budget" approach taught in most managerial accounting textbooks (e.g., Datar and Rajan 2021; Garrison, Noreen, and Brewer 2017; Farmer and Fredin 2022). With this approach, granular forecasts for each cost component are linearly aggregated to form higher-level forecasts. While this bottom-up process is intuitive and widely adopted, our theory suggests it leads to underestimation when (i) costs exhibit right-skew, as is typical, and (ii) some of the component estimates reflect modes or medians. Thus, the limitations of this approach as well as the conditions necessary for linear aggregation to be unbiased should be more clearly outlined (i.e., either component distributions are symmetric, on average, or all component estimates being summed are means – the only statistic that is linear).

From a practical perspective, organizations may benefit from carefully reviewing their forecasting systems with aggregation fallacy in mind. While much attention has been devoted to biases in estimating individual probability distributions (Griffin and Tversky 1992; Sharot, Korn, and Dolan 2011; Tversky and Kahneman 1974), we emphasize the need to consider how those individual estimates are aggregated – to which organizations may currently devote less attention. A lack of attention in this area can be problematic because, as we demonstrate, individuals do not aggregate optimally, with many erroneously assuming probabilistic information can be linearly aggregated like deterministic outcomes. The widespread use of spreadsheets further exacerbates the issue, making it unlikely that managers recognize and correct for errors in the aggregation of forecasts. Thus, organizations may benefit from investing in training programs to raise awareness of aggregation fallacy or formal controls to enforce scrutiny of aggregation policies.

Our findings also suggest that the use of technology in forecasting might be more beneficial than previously thought. When sufficient historical information is available, future costs can be predicted using algorithms that preclude errors in the aggregation of forecasts. However, many organizations continue to rely on human judgment (Bean 2023; Brau, Aloysius, and Siemsen 2023; Goodwin et al. 2023; Liu 2022), and technology can be limited for one-off capital budgeting projects or new ventures with limited historical data. Business conditions can also change rapidly, quickly rendering historical data obsolete. Another approach is to combine technology with human judgment. Employees can use their domain expertise to subjectively estimate probability distributions for each cost component, and then algorithms can be designed to aggregate this information optimally – beyond the simple addition used in most spreadsheets. The use of triangular distributions in practice makes this approach feasible, as they only require managers to estimate a minimum, maximum, and mode (which can be obtained with relative ease compared to other statistics), while flexibly allowing for skew.

Our study opens the door for future research on potential solutions to aggregation fallacy. In our experiments, we find mixed evidence for one such solution. That is, making the midpoint of the aggregate distribution more salient reduces aggregation fallacy in some cases, because the aggregate midpoint acts as a reference point, such that individuals may question the summing of modes or medians. Another approach that can be addressed by future research is to modify the forecasting task such that managers do not aggregate. For example, as outlined above, managers may forecast component costs while leaving aggregation to an algorithm. Aggregation may also be minimized by making forecasting less granular. Instead of asking managers to forecast and aggregate each component of cost (e.g., cost of each project stage), they could be asked to directly forecast costs at broader levels (e.g., total project costs). In the latter case, managers might use an "outside view," forming estimates by considering total costs from similar situations in the past, thereby avoiding explicit aggregation (Lovallo and Kahneman 2003).

Finally, our study should be considered within the broader scope of budgeting as a tool for both planning and control. Our focus has been on the planning function, where forecast accuracy is the primary concern. However, when budgets are used as targets for performance evaluation, organizations may prefer cost underestimation (harder targets) or overestimation (easier targets). Thus, evaluating whether the underestimation that results from aggregation fallacy is undesirable becomes more complex. In fact, when budgets are used in performance evaluations, employees may have an incentive to inflate cost forecasts to obtain easier targets. Thus, the underestimation resulting from aggregation fallacy could counteract overestimation that results from incentives to obtain easier targets.

REFERENCES

Adams, K. 2024. Twelve common startup budgeting mistakes to avoid. *Burkland* (August 6). Available at https://burklandassociates.com/2024/08/06/12-common-startup-budgeting-mistakes-to-avoid/.

Amar, R. 2025. Why construction cost overruns follow a fat-tailed distribution (and why it matters!). *Medium* (March 6). Available at https://medium.com/ladder-ai/why-construction-cost-overruns-follow-a-fat-tailed-distribution-and-why-it-matters-29e639741624.

Andrews, D. W. K. 1991. An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis* 38 (2): 187-203.

Banks, J., J. S. Carson, B. L. Nelson, and D. M. Nicol. 2004. *Discrete-Event System Simulation.* Fourth Edition, London: Pearson Prentice Hall.

Barber, J. A., and S. G. Thompson. 1998. Analysis and interpretation of cost data in randomised controlled trials: Review of published studies. *British Medical Journal* 317 (7167): 1195-1200.

Barefield, R. M. 1969. Comments on a measure of forecasting performance. *Journal of Accounting Research* 7 (2): 324-327.

Basu, S., and S. Markov. 2004. Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics* 38: 171-203.

Bean, R. 2023. Has progress on data, analytics, and AI stalled at your company? *Harvard Business Review* (January 30). Available at https://hbr.org/2023/01/has-progress-on-data-analytics-and-ai-stalled-at-your-company.

Berry, A., and D. Otley. 1975. The aggregation of estimates in hierarchical organizations. *Journal of Management Studies* 12 (1-2): 175-193.

Book, S. A. 2001. Estimating probably system cost. *Crosslink* 2 (1): 12-21.

Brau, R., J. Aloysius, and E. Siemsen. 2023. Demand planning for the digital supply chain: How to integrate human judgment and predictive analytics. *Journal of Operations Management* 69 (6): 965-982.

Briggs, A., and A. Gray. 1998. The distribution of health care costs and their statistical analysis for economic evaluation. *Journal of Health Services Research & Policy* 3 (4): 233-245.

Brüggen, A., and J. Luft. 2011. Capital rationing, competition, and misrepresentation in budget forecasts. *Accounting, Organizations and Society* 36 (7): 399-411.

Brüggen, A., and J. L. Luft. 2016. Cost estimates, cost overruns, and project continuation decisions. *The Accounting Review* 91 (3): 793-810.

Chen, W., T. Kravet, and B. Ren. 2025. Forecasting the median: Revisiting the puzzling parity in management and analyst forecast accuracy. University of Connecticut (Working Paper). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5284497.

Christensen, J. 2010. Accounting errors and errors of accounting. *The Accounting Review* 85 (6): 1827-1838.

Datar, S. M., and M. V. Rajan. 2021. *Horngren's Cost Accounting: A Managerial Emphasis*. 17th, Global Edition ed. Harlow: Pearson Education Limited.

de Langhe, B., S. Puntoni, and R. P. Larrick. 2017. Linear thinking in a nonlinear world. *Harvard Business Review* 95 (3): 130-139.

Elliott, R., and K. Payne. 2005. *Essentials of Economic Evaluation in Healthcare.* London, UK: Pharmaceutical Press.

Engelberg, J., C. F. Manski, and J. Williams. 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics* 27 (1): 30-41.

Farmer, K. C., and A. Fredin. 2022. *Cost Accounting: With Integrated Data Analytics*. 1st ed. Hoboken, NJ: Wiley.

Flyvbjerg, B., A. Ansar, A. Budzier, S. Buhl, C. Cantarelli, M. Garbuio, C. Glenting, M. S. Holm, D. Lovallo, and D. Lunn. 2018. Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice* 118: 174-190.

Flyvbjerg, B., M. S. Holm, and S. Buhl. 2002. Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association* 68 (3): 279-295.

Flyvbjerg, B., M. K. Skamris Holm, and S. L. Buhl. 2003. How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews* 23 (1): 71-88.

Garrison, R. H., E. W. Noreen, and P. C. Brewer. 2018. *Managerial Accounting*. 16th ed. New York, NY: McGraw-Hill Education.

Goodwin, P., J. Hoover, S. Makridakis, F. Petropoulos, and L. Tashman. 2023. Business forecasting methods: Impressive advances, lagging implementation. *PLOS One* 18 (12): e0295693.

Gray, D. A. 2024. Why data science projects fail: Part 8. Principles for Successful Analytics Projects (February 8). Available at https://doi.org/10.1287/LYTX.2024.01.10.

Griffin, D., and A. Tversky. 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24 (3): 411-435.

Gu, Z., and J. S. Wu. 2003. Earnings skewness and analyst forecast bias. *Journal of Accounting and Economics* 35 (1): 5-29.

Haka, S. F. 2006. A review of the literature on capital budgeting and investment appraisal: Past, present, and future musings. *Handbooks of Management Accounting Research* 2: 697-728.

Hogg, R. V., J. McKean, and A. T. Craig. 2012. *Introduction to Mathematical Statistics*. Upper Saddle River, NJ: Pearson Education.

Jackson, S. B., L. Salzsieder, and T. J. Schaefer. 2025. Do managers pursue their budget goals using revenues or expenses? *The Accounting Review* 100 (3): 251-276.

Johnson, D. 1997. The triangular distribution as a proxy for the beta distribution in risk analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (3): 387-398.

Jørgensen, M. 2014. Communication of software cost estimates. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* 28: 1-5.

Juslin, P., M. Lindskog, and B. Mayerhofer. 2015. Is there something special with probabilities?–Insight vs. computational ability in multiple risk combination. *Cognition* 136: 282-303.

Lovallo, D., and D. Kahneman. 2003. Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review* 81 (7): 56-63.

Kaplan, R. S., and S. R. Anderson. 2007. *Time-Driven Activity-Based Costing: A Simpler and More Powerful Path to Higher Profits*. Boston, MA: Harvard Business School Press.

Karlsson, M., Y. Wang, and N. R. Ziebarth. 2024. Getting the right tail right: Modeling tails of health expenditure distributions. *Journal of Health Economics* 97: 102912.

Klakegg, O. J., and S. Lichtenberg. 2016. Successive cost estimation-successful budgeting of major projects. *Procedia – Social and Behavioral Sciences* 226: 176-183.

Labro, E. 2025. Thoughts on management accounting research at The Accounting Review's Centennial. *The Accounting Review*: 1-12. Available at https://doi.org/10.2308/TAR-2025-0481.

Larrick, R. P., and J. B. Soll. 2008. The MPG illusion. *Science* 320 (5883): 1593-1594.

Liu, M. 2022. Assessing human information processing in lending decisions: A machine learning approach. *Journal of Accounting Research* 60 (2): 607-651.

Lock, R. H., P. F. Lock, K. L. Morgan, E. F. Lock, and D. F. Lock. 2021. *Statistics: Unlocking the Power of Data*. Third Edition. John Wiley & Sons: Hoboken, New Jersey.

Love, P. E. D., J. Smith, I. Simpson, M. Regan, and O. Olatunji. 2015. Understanding the landscape of overruns in transport infrastructure projects. *Environment and Planning B: Planning and Design* 42 (3): 490-509.

Manju, M. A., M. J. Candel, and G. J. Van Breukelen. 2021. Robustness of cost-effectiveness analyses of cluster randomized trials assuming bivariate normality against skewed cost data. *Computational Statistics & Data Analysis* 157: 107143.

Merrow, E. W. 2024. *Industrial Megaprojects: Concepts, Strategies, and Practices for Success*: John Wiley & Sons, Hoboken, New Jersey.

Mohan, S., M. Gopalakrishnan, H. Balasubramanian, and A. Chandrashekar. 2007. A lognormal approximation of activity duration in PERT using two time estimates. *Journal of the Operational Research Society* 58 (6): 827-831.

Mun, J. 2006. *Modeling risk: Applying Monte Carlo simulation, real options analysis, forecasting, and optimization techniques*. Vol. 347: John Wiley & Sons.

Otley, D. T. 1985. The accuracy of budgetary estimates: Some statistical evidence. *Journal of Business Finance & Accounting* 12 (3): 415-428.

Peterson, C., and A. Miller. 1964. Mode, median, and mean as optimal strategies. *Journal of Experimental Psychology* 68 (4): 363-367.

Schwartz, S. T., E. E. Spires, D. E. Wallin, and R. A. Young. 2012. Aggregation in budgeting: An experiment. *Journal of Management Accounting Research* 24 (1): 177-199.

Sharot, T., C. W. Korn, and R. J. Dolan. 2011. How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience* 14 (11): 1475-1479.

Stango, V., and J. Zinman. 2009. Exponential growth bias and household finance. *Journal of Finance* 64 (6): 2807-2849.

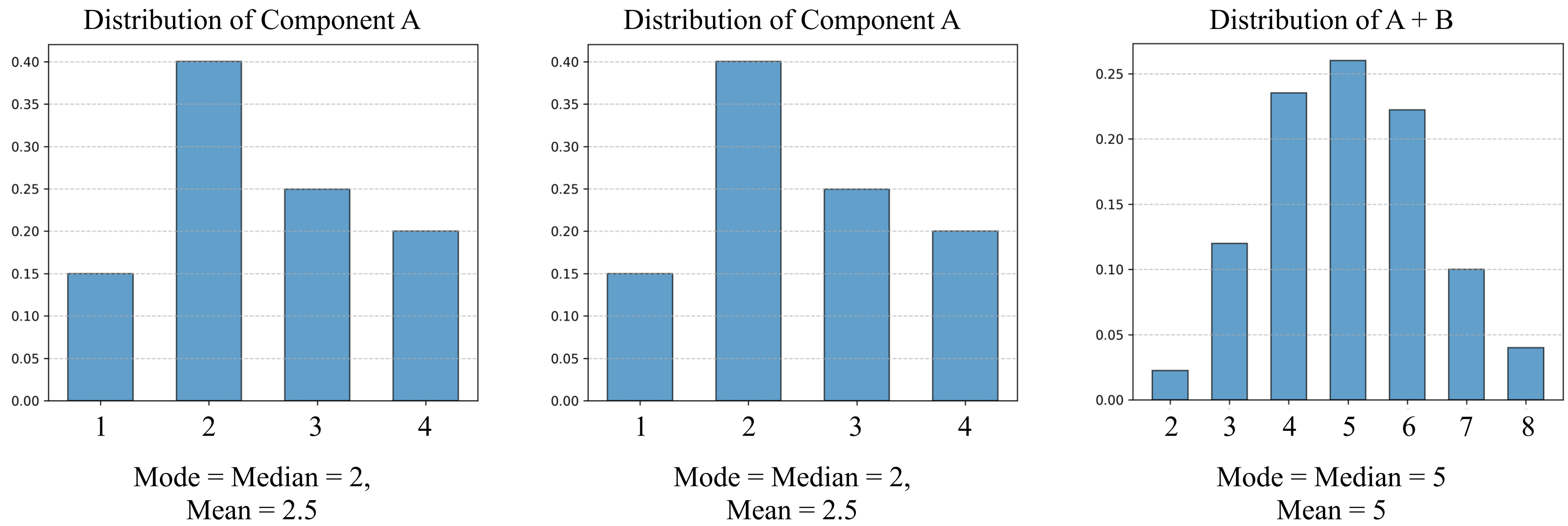
Turvey, R. 1971. Discussion of 'The management of capital investment.' *Journal of the Royal Statistical Society, Series A (General)* 134 (4): 529.

Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 (4157): 1124-1131.

Udoumoh, E. F., and D. W. Ebong. 2017. A review of activity time distributions in risk analysis. American Journal of Operations Research 7 (6): 356-371.

Zhang, X., O. L. Olvera Astivia, E. Kroc, and B. D. Zumbo. 2023. How to think clearly about the central limit theorem. *Psychological Methods* 28 (6): 1427-1445.

Figure 1: Example with Right-Skew Component Distributions

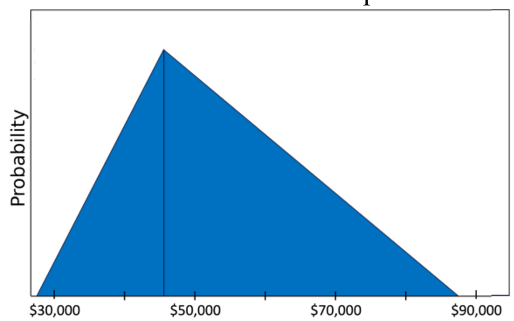


This figure illustrates an example, demonstrating the dynamics of the Central Limit Theorem. It shows the distributions of two independent and identically distributed components with right skew, A and B (left and middle), and the distribution of their sum, A + B (right).

Figure 2: Experiments 1a and 1b – Distributions of Component and Project Costs

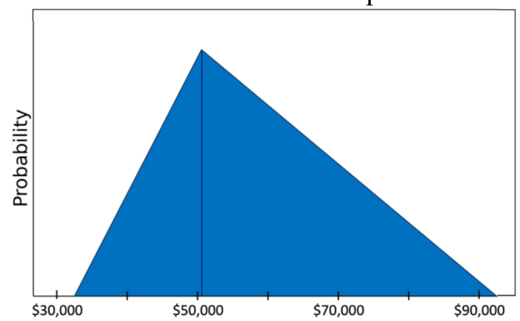
Panel A: Project with two right-skewed cost components

Distribution of Cost Component A



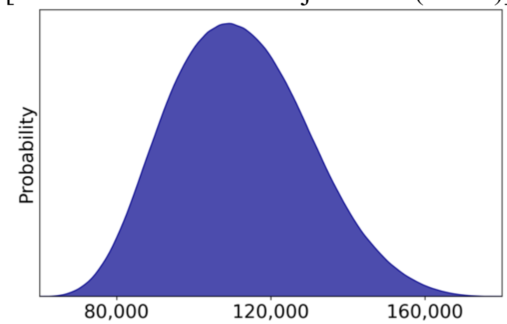
Min: $27,500; Mode: $45,500; Max: $87,500
[Median: $52,000; Mean: $53,500; Mid: $57,500]

Distribution of Cost Component B



Min: $32,500; Mode: $50,500; Max: $92,500
[Median: $57,000; Mean: $58,500; Mid: $62,500]

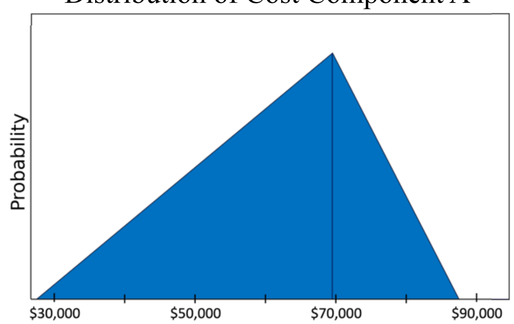
[Distribution of Total Project Cost (A + B)]



[Min: $60,000; Mode: $110,237; Max: $180,000]
[Median: $111,130; Mean: $112,000; Mid: $120,000]

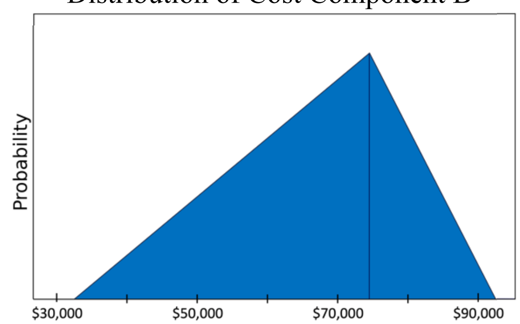
Panel B: Project with two left-skewed cost components

Distribution of Cost Component A



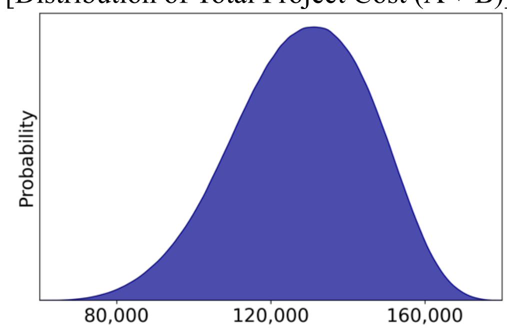
Min: $27,500; Mode: $69,500; Max: $87,500
[Median: $63,000; Mean: $61,500; Mid: $57,500]

Distribution of Cost Component B



Min: $32,500; Mode: $74,500; Max: $92,500
[Median: $68,000; Mean: $66,500; Mid: $62,500]

[Distribution of Total Project Cost (A + B)]

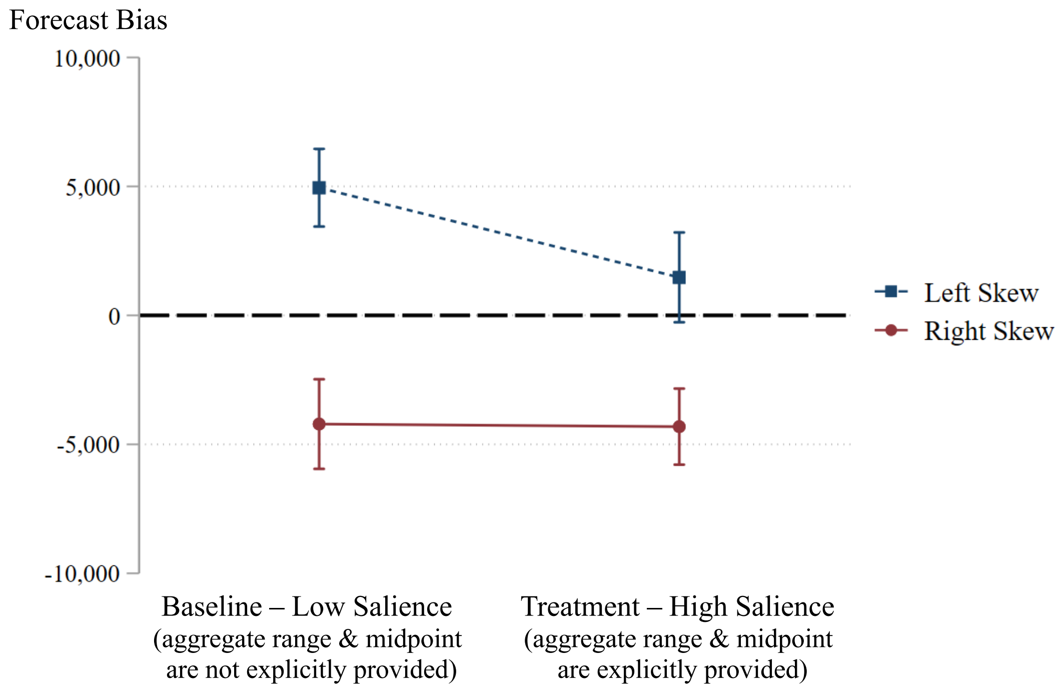


[Min: $60,000; Mode: $130,091; Max: $180,000]
[Median: $128,870; Mean: $128,000; Mid: $120,000]

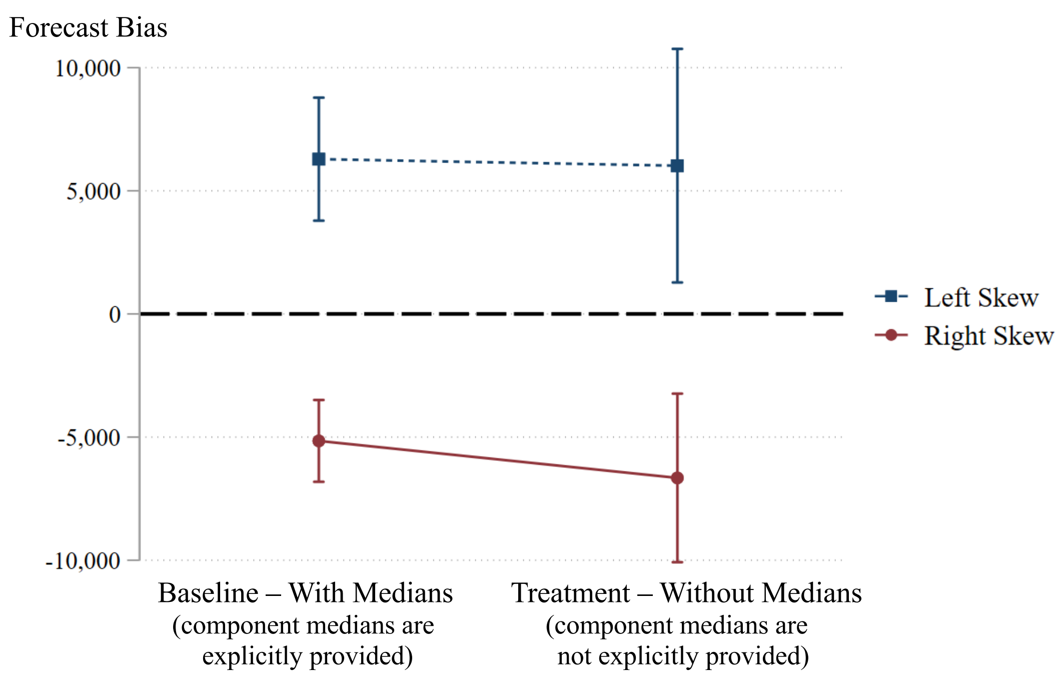
This figure illustrates the distributions used in Experiments 1a and 1b. In each panel, the first / second / third graph from left to right shows the distribution of cost component A / cost component B / total project cost (A + B). Panel A shows the distributions for the project with right-skewed components, while Panel B shows the distributions for the project with left-skewed components. Note that participants do not observe the distribution of the total project cost.

Figure 3: Experiments 1a and 1b – Forecast Bias

Panel A: Experiment 1a – Forecast Bias (Forecasted Total Cost – Mean of Total Cost)

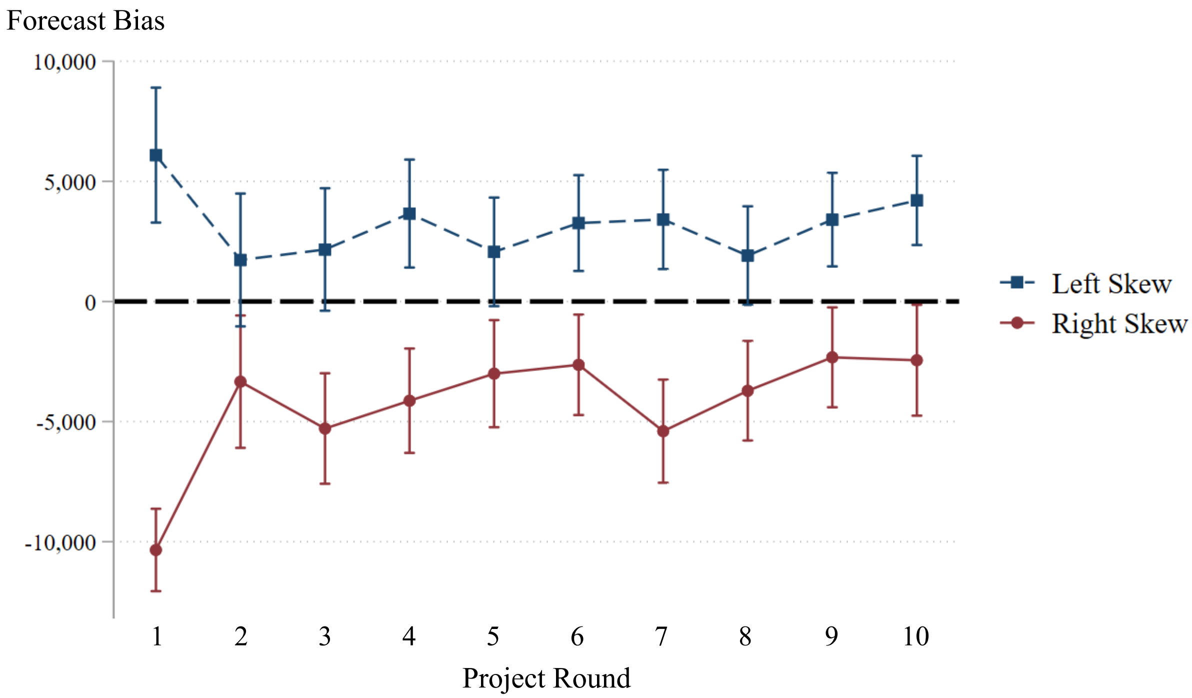


Panel B: Experiment 1b – Forecast Bias (Forecasted Total Cost – Mean of Total Cost)



This figure shows forecast bias – with 90% confidence intervals – in each of the four conditions in Experiment 1a (Panel A) and Experiment 1b (Panel B).

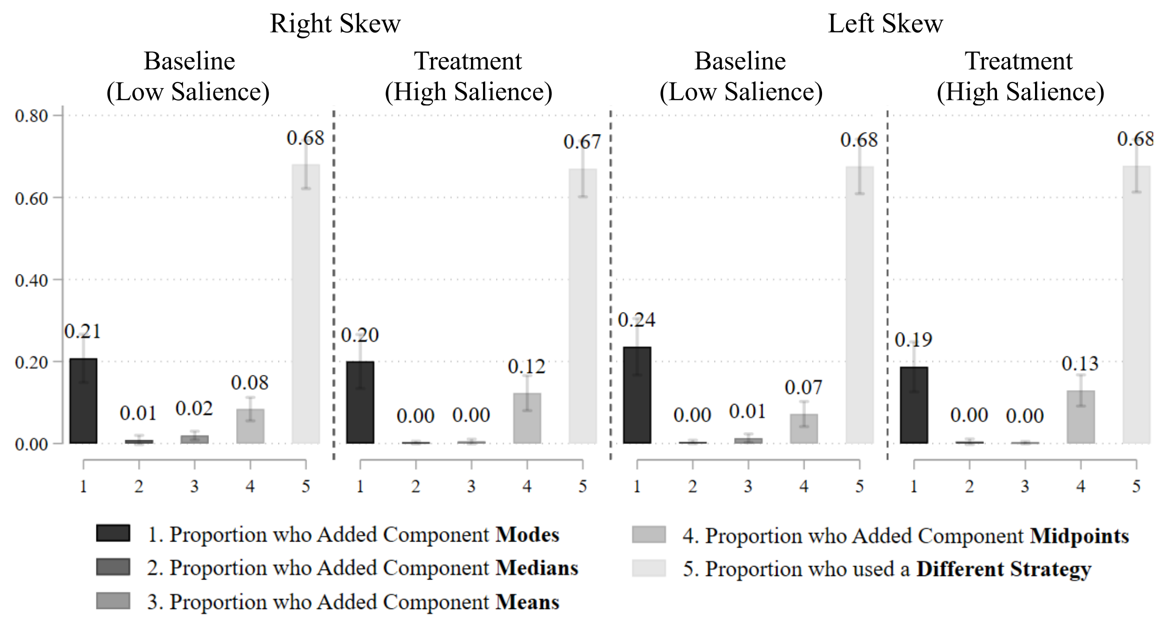
Figure 4: Experiment 1a – Forecast Bias over Time



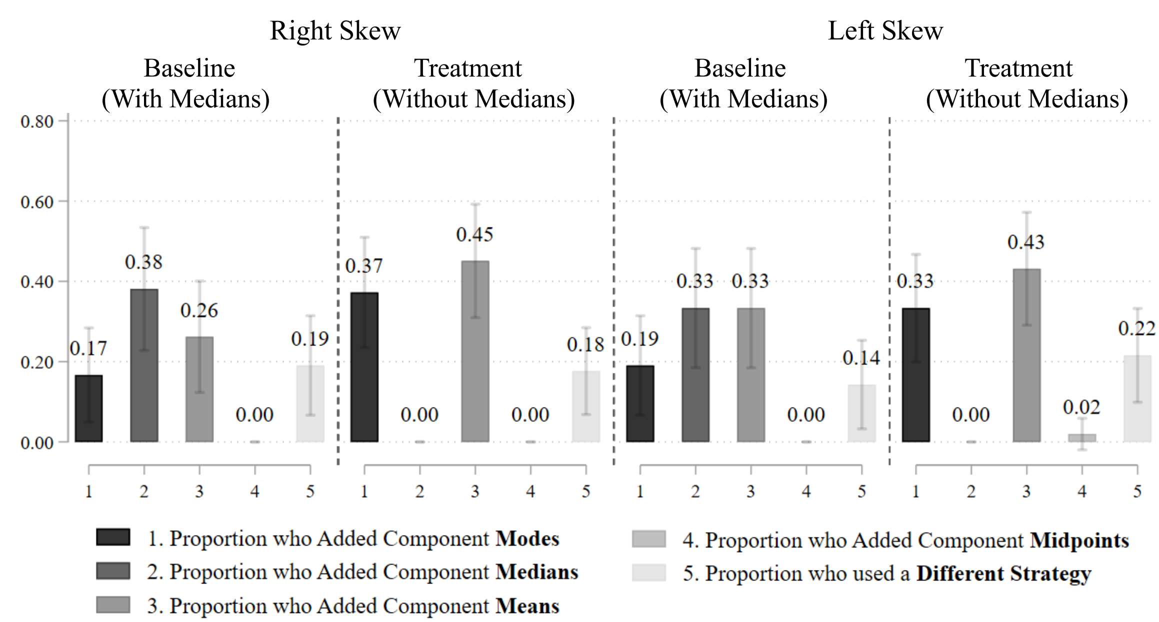
This figure shows forecast bias – with 90% confidence intervals – across the ten project rounds in Experiment 1a for projects with right-skewed cost components (with red circles) and projects with left-skewed cost components (with blue squares) – pooled across the baseline and treatment conditions.

Figure 5: Experiments 1a and 1b – Proportion of Forecasts Reflecting Various Strategies

Panel A: Experiment 1a



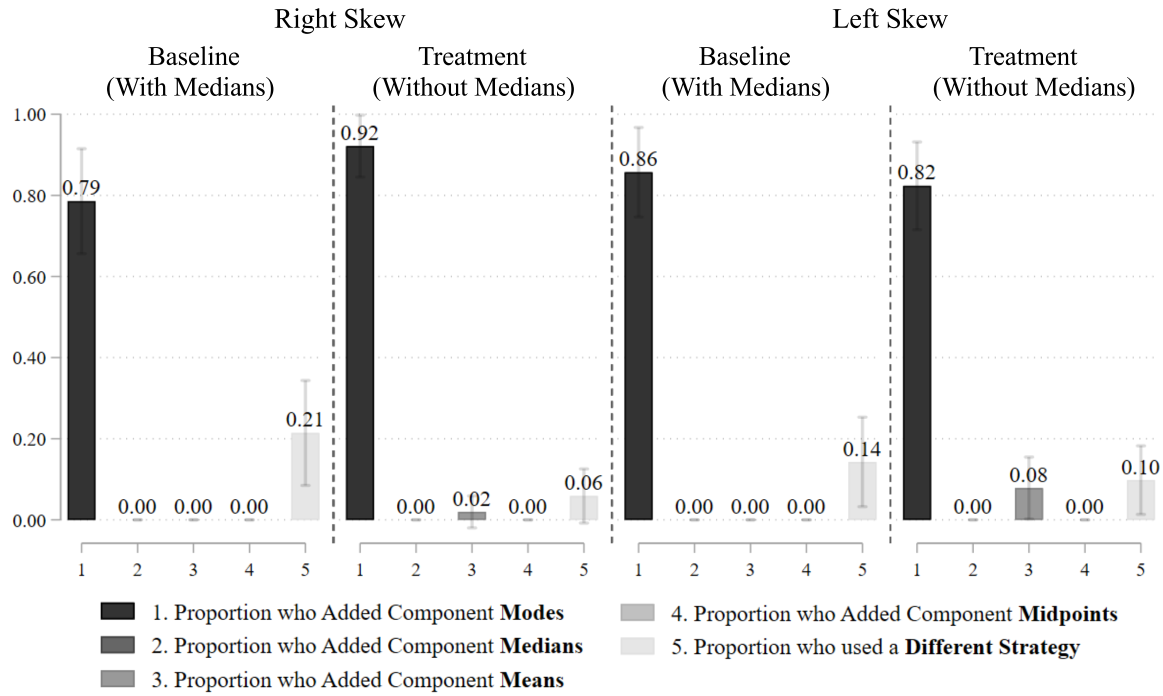
Panel B: Experiment 1b



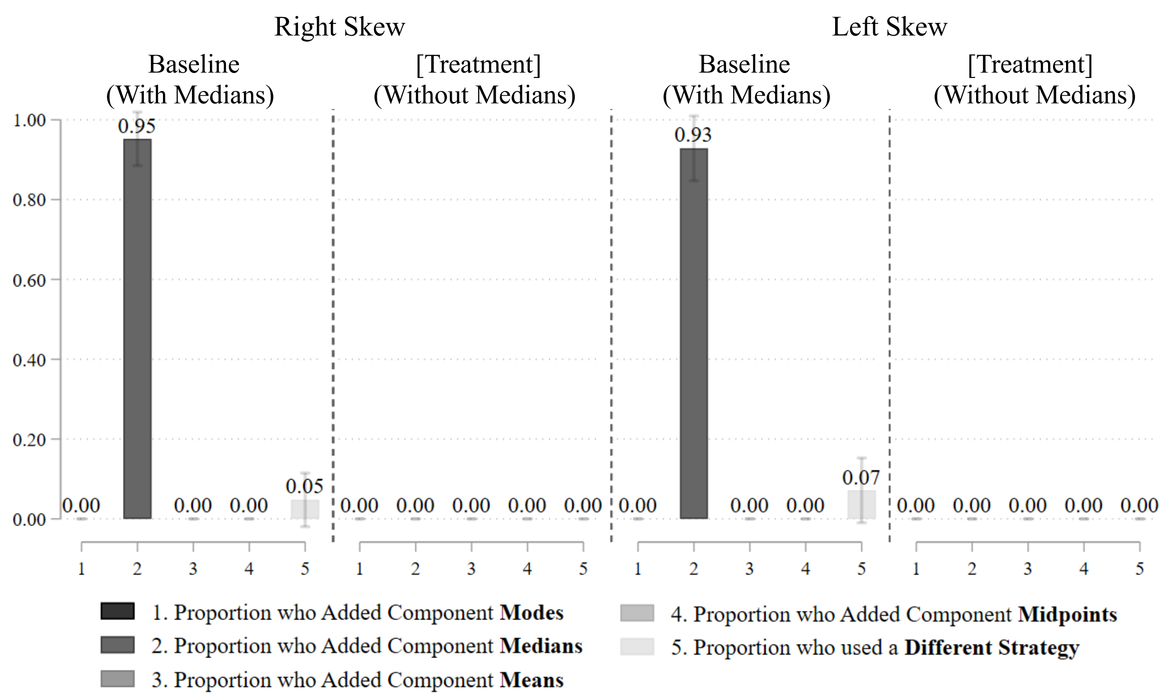
This figure depicts the proportions of forecasts reflecting various forecasting strategies, including the proportions of forecasts that equal the sum of the (1) component modes, (2) medians, (3) means, and (4) midpoints, as well as (5) the proportion of forecasts that do no fall in any of those categories (i.e., different strategy) separately for each of the four conditions in both Experiment 1a (Panel A) and Experiment 1b (Panel B). See the Methods section for full information on the baseline and treatment conditions as well as the means, modes, medians, and midpoints of the distributions for the cost components and total cost.

Figure 6: Experiment 1b – Additional Forecasts

Panel A: Forecasts of Aggregate Mode

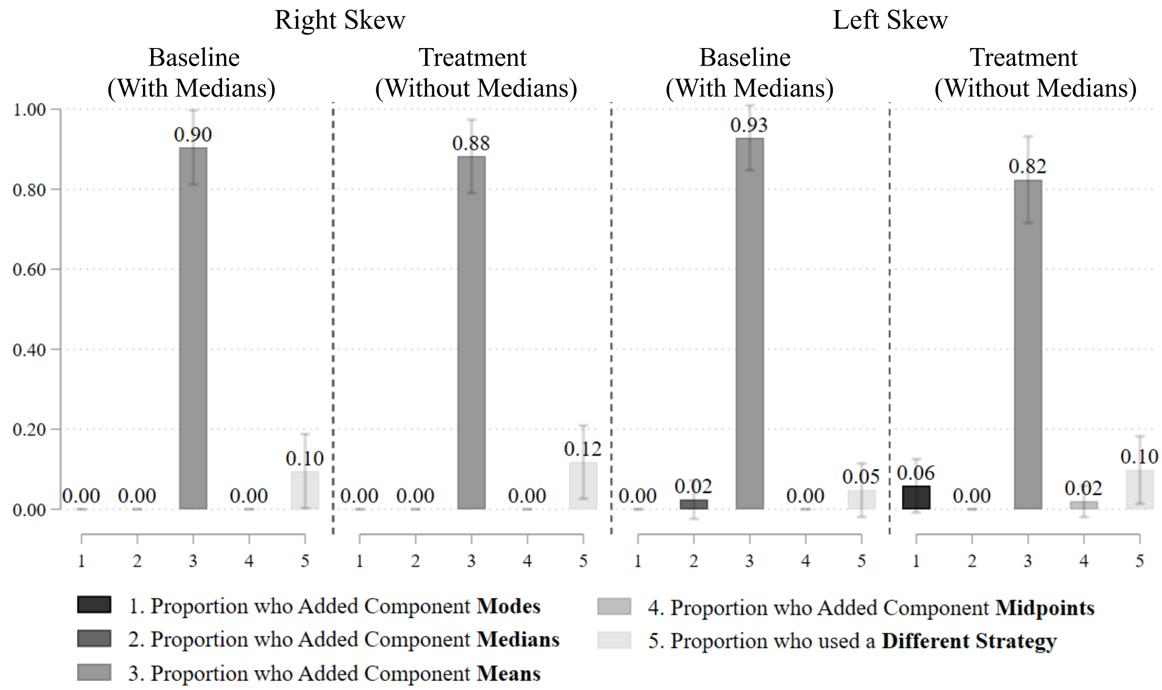


Panel B: Forecasts of Aggregate Median (Baseline Only)



(Figure continues on next page)

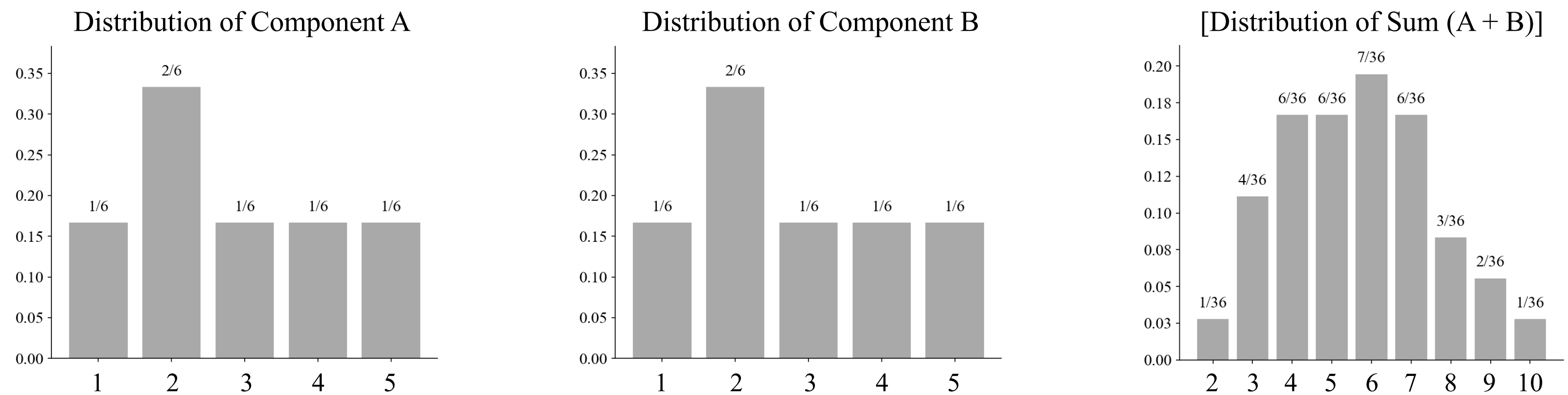
Panel C: Forecasts of Aggregate Mean



This figure depicts the proportions of forecasts reflecting various forecasting strategies, including the proportions of forecasts that equal the sum of the (1) component modes, (2) medians, (3) means, and (4) midpoints, as well as (5) the proportion of forecasts that do no fall in any of those categories (i.e., different strategy) separately for each of the four conditions, when participants forecast the aggregate mode (in Panel A), aggregate median (in Panel B, baseline conditions only), and aggregate mean (in Panel C). See the Methods section for full information on the baseline and treatment conditions as well as the means, modes, medians, and midpoints of the distributions for the cost components and total cost.

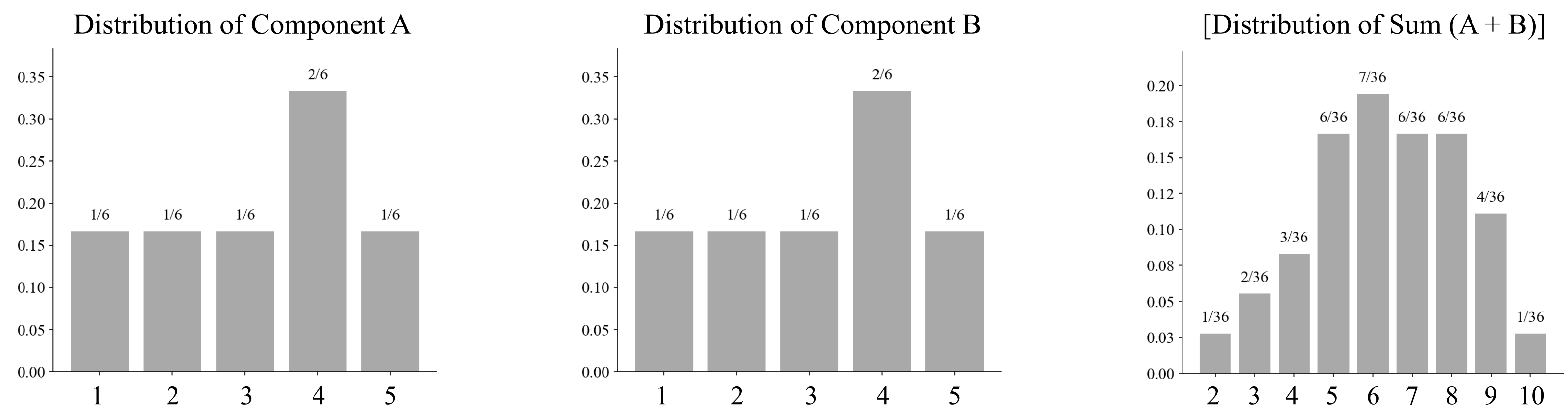
Figure 7: Experiments 2a, 2b, and 2c – Distributions of Components and Sum

Panel A: Task with two right-skewed components



Distribution of Component A	Distribution of Component B	[Distribution of Sum (A + B)]
Mode: 2; Median: 2.5; Mean: 2.83; Midpoint: 3	Mode: 2; Median: 2.5; Mean: 2.83; Midpoint: 3	Mode: 6; Median: 6; Mean: 5.67; Midpoint: 6

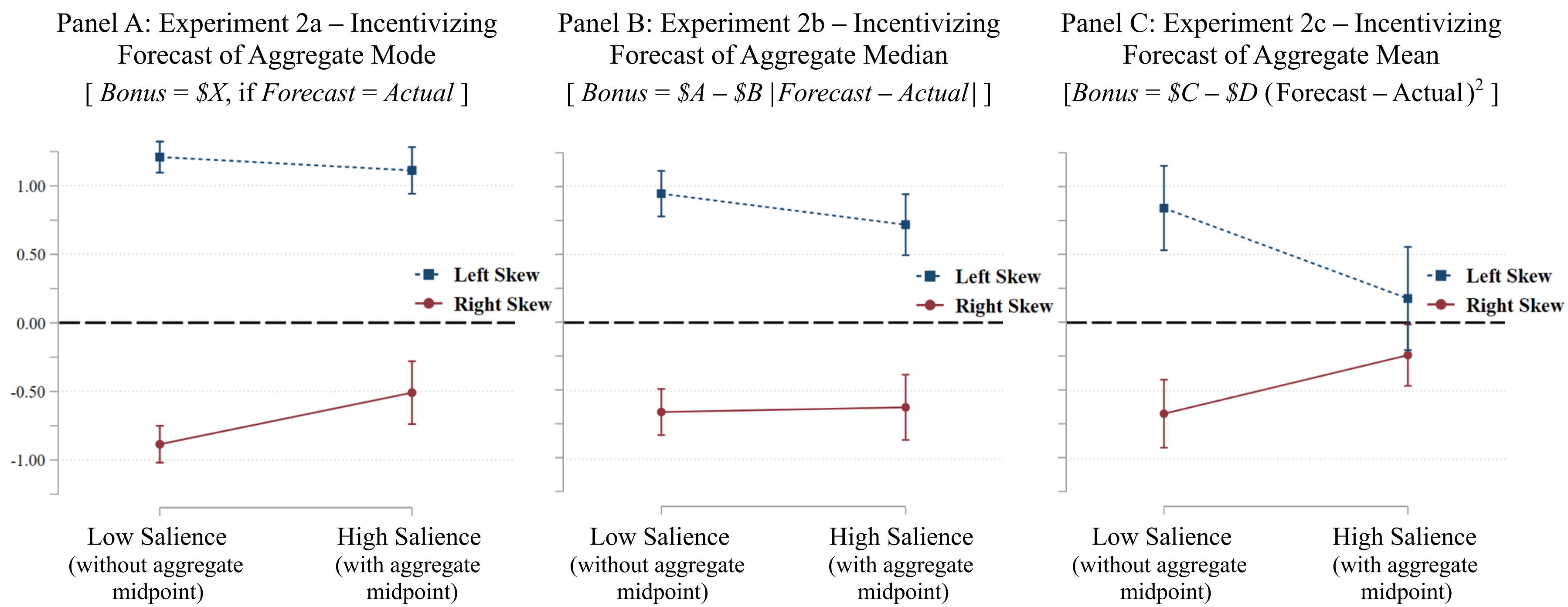
Panel B: Task with two left-skewed components



Distribution of Component A	Distribution of Component B	[Distribution of Sum (A + B)]
Mode: 4; Median: 3.5; Mean: 3.17; Midpoint: 3	Mode: 4; Median: 3.5; Mean: 3.17; Midpoint: 3	Mode: 6; Median: 6; Mean: 6.33; Midpoint: 6

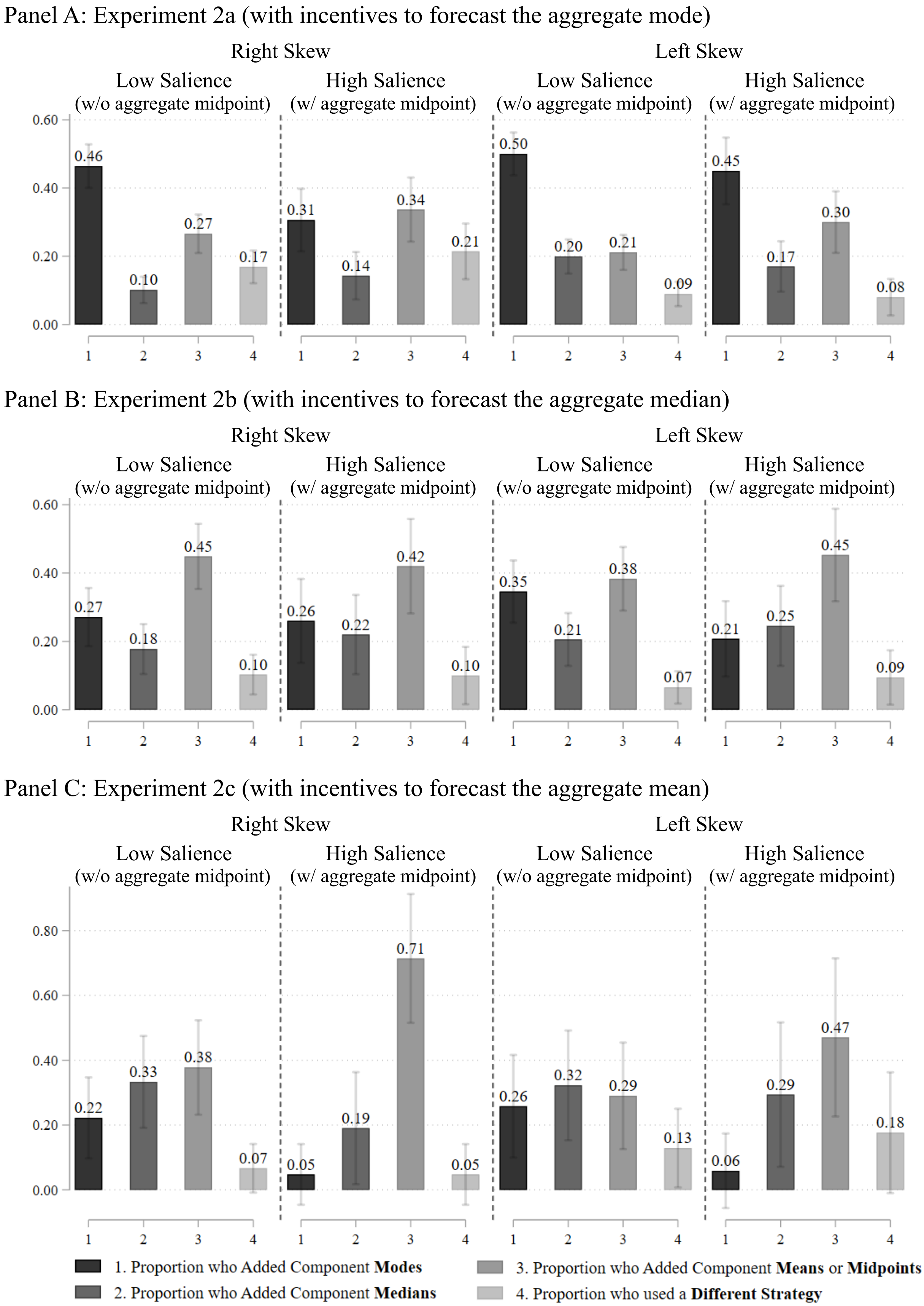
This figure shows the distribution of the components (A and B) in the top row and the distribution of the sum of the components (A + B) in the bottom row for experiment 2. The right-skewed components are shown in the left column and the left-skewed components are shown in the right column. Recall that within each skew condition, the two components being summed have identical distributions.

Figure 8: Experiments 2a, 2b, and 2c – Forecast Bias



This figure shows forecast bias – measured relative to the optimal forecast of 6 – with 90% confidence intervals in each of the four conditions in Experiment 2a (Panel A), Experiment 2b (Panel B), and Experiment 2c (Panel C). Experiments 2a, 2b, and 2c differ to the extent that they incentivize forecasting the aggregate mode, median, and mean respectively. In each experiment, it is optimal for participants to provide a forecast of 6, irrespective of the experimental condition.

Figure 9: Experiments 2a, 2b, and 2c – Proportion of Forecasts Reflecting Various Strategies



This figure depicts the proportions of forecasts reflecting various forecasting strategies, including the proportions of forecasts that equal the sum of the (1) component modes, (2) medians, (3) means, and (4) midpoints, as well as (5) the proportion of forecasts that do no fall in any of those categories (i.e., different strategy) separately for each of the four conditions in Experiment 2a (Panel A), Experiment 2b (Panel B), and Experiment 2c (Panel C).

Table 1: Experiments 1a and 1b – Descriptive Statistics

Panel A: Experiment 1a – Forecast Bias [N = 3,460]

	Baseline – Low Salience (aggregate range & midpoint are not explicitly provided)	Treatment – High Salience (aggregate range & midpoint are explicitly provided)	
Right Skew	-4,213 (18,065) [N = 930]	-4,313 (17,445) [N = 830]	Δ = -100
Left Skew	4,948 (16,625) [N = 780]	1,474 (19,279) [N = 920]	Δ = -3,474
	Δ = 9,161	Δ = 5,787	Δ_{Δ} = 3,373

Panel B: Experiment 1b – Forecast Bias [N = 186]

	Baseline – With Medians (component medians are explicitly provided)	Treatment – Without Medians (component medians are not explicitly provided)	
Right Skew	-5,155 (6,462) [N = 42]	-6,657 (14,653) [N = 51]	Δ = -1,502
Left Skew	6,286 (9,726) [N = 42]	6,020 (20,303) [N = 51]	Δ = -266
	Δ = 11,440	Δ = 12,676	Δ_{Δ} = -1,236

This table shows average forecast bias (forecast – aggregate mean) for each of the four conditions in Experiment 1a (Panel A) and Experiment 1b (Panel B). Standard deviations are provided in parentheses and the number of forecasts is depicted in brackets (in Experiment 1a, each participant provides ten forecasts; in Experiment 1b, each participant provides two forecasts, i.e., one for each skew).

Table 2: Experiments 1a and 1b – Regression Results

Panel A: Experiment 1a – Regression Results

	(1) *Forecast Bias* (relative to mean, all ACs correct)	(2) *Forecast Bias* (relative to median, all ACs correct)	(3) *Forecast Bias* (relative to mean, full sample)
Left Skew [β_1]	9,161***	7,421***	5,751***
	(1,392)	(1,392)	(1,606)
High Salience [β_2]	-100	-100	100
	(1,381)	(1,381)	(1,294)
Left Skew x High Salience [β_3]	-3,373*	-3,373*	-182
	(1,962)	(1,962)	(2,082)
Constant [β_0]	-4,213***	-3,343***	-4,360***
	(1,053)	(1,053)	(992)
# of Forecasts	3,460	3,460	5,040
# of Clusters (i.e., Participants)	346	346	504
R^2	0.04	0.03	0.02
Remaining Simple Effects			
Left Skew when *High Salience* = 1 [$\beta_1 + \beta_3$]	5,787***	4,047***	5,569***
	(1,383)	(1,383)	(1,325)
High Salience when *Left Skew* =1 [$\beta_2 + \beta_3$]	-3,474**	-3,474**	-82
	(1,394)	(1,394)	(1,631)
Forecast Bias in each Condition			
(1) Right Skew and Low Salience [β_0]	-4,213***	-3,343***	-4,360***
	(1,053)	(1,053)	(992)
(2) Left Skew and Low Salience [$\beta_0 + \beta_1$]	4,948***	4,078***	1,392
	(911)	(911)	(1,263)
(3) Right Skew and High Salience [$\beta_0 + \beta_2$]	-4,313***	-3,443***	-4,260***
	(894)	(894)	(832)
(4) Left Skew and High Salience [$\beta_0 + \beta_1 + \beta_2 + \beta_3$]	1,474	604	1,310
	(1,055)	(1,055)	(1,031)

Panel B: Experiment 1b – Regression Results

	(1) *Forecast Bias* (relative to mean, all ACs correct)	(2) *Forecast Bias* (relative to median, all ACs correct)	(3) *Forecast Bias* (relative to mean, full sample)
Left Skew [β_1]	11,440***	9,700***	10,254***
	(1,997)	(1,997)	(1,641)
No Median [β_2]	-1,502	-1,502	-1,146
	(2,289)	(2,289)	(1,838)
Left Skew x No Median [β_3]	1,236	1,236	299
	(3,680)	(3,680)	(3,243)
Constant [β_0]	-5,155***	-4,285***	-5,346***
	(999)	(999)	(806)
# of Forecasts	186	186	262
# of Clusters (i.e., Participants)	93	93	131
R^2	0.16	0.12	0.13
Remaining Simple Effects			
Left Skew when *No Median* = 1 [β_1 + β_3]	12,676***	10,936***	10,553***
	(3,091)	(3,091)	(2,798)
No Median when *Left Skew* =1 [β_2 + β_3]	-266	-266	-847
	(3,225)	(3,225)	(2,856)
Forecast Bias in each Condition			
(1) Right Skew and No Median [β_0]	-5,155***	-4,285***	-5,346***
	(999)	(999)	(806)
(2) Left Skew and No Median [β_0 + β_1]	6,286***	5,416***	4,908***
	(1,503)	(1,503)	(1,239)
(3) Right Skew and Median [β_0 + β_2]	-6,657***	-5,787***	-6,492***
	(2,059)	(2,059)	(1,652)
(4) Left Skew and Median [β_0 + β_1 + β_2+ β_3]	6,020**	5,150*	4,061
	(2,853)	(2,853)	(2,574)

This table shows the results of regressing *Forecast Bias* on indicator variables capturing the experimental manipulations and their interaction term for Experiment 1a (Panel A) and for Experiment 1b (Panel B). *Left Skew* equals zero (one) for projects with right-skewed (left-skewed) component distributions. *High Salience* equals zero when aggregate range and midpoint are not explicitly provided and one when they are explicitly provided (i.e., highly salient). *No Median* equals zero when the component medians are not explicitly provided – in addition to component modes and means – and one when component medians are explicitly provided. Columns 1 and 2 ("all ACs correct") limit the sample to participants, who correctly answer all 11 attention checks [ACs]. Column 3 ("full sample") reflect the entire sample of participants. Column 2 ("relative to median") shows the results when forecast bias is derived relative to the aggregate median rather than the aggregate mean (full details are provided in the main text).

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, referring to two-tailed tests. Standard errors are clustered within participants (see # of Clusters) and presented in parentheses.

Table 3: Experiments 2a, 2b, and 2c – Descriptive Statistics

Panel A: Experiment 2a (Incentivizing Mode) – Forecast Bias [N = 681]

	Low Salience (without aggregate midpoint)	High Salience (with aggregate midpoint)	
Right Skew	-0.89 (1.25) [N = 237]	-0.51 (1.38) [N = 98]	Δ = 0.38
Left Skew	1.21 (1.08) [N = 246]	1.11 (1.03) [N = 100]	Δ = -0.10
	Δ = 2.09	Δ = 1.62	Δ_Δ = -0.47

Panel B: Experiment 2b (Incentivizing Median) – Forecast Bias [N = 317]

	Low Salience (without aggregate midpoint)	High Salience (with aggregate midpoint)	
Right Skew	-0.65 (1.06) [N = 107]	-0.62 (1.03) [N = 50]	Δ = 0.03
Left Skew	0.94 (1.05) [N = 107]	0.72 (0.99) [N = 53]	Δ = -0.23
	Δ = 1.60	Δ = 1.34	Δ_Δ = -0.26

Panel C: Experiment 2c (Incentivizing Mean) – Forecast Bias [N = 114]

	Low Salience (without aggregate midpoint)	High Salience (with aggregate midpoint)	
Right Skew	-0.67 (1.00) [N = 45]	-0.24 (0.63) [N = 21]	Δ = 0.43
Left Skew	0.84 (1.04) [N = 31]	0.18 (0.95) [N = 17]	Δ = -0.66
	Δ = 1.51	Δ = 0.41	Δ_Δ = -1.09

This table presents average forecast bias (calculated as: forecast – aggregate benchmark of six) for each of the four conditions in Experiment 2a (Panel A), Experiment 2b (Panel B), and Experiment 2c (Panel C). Standard deviations are provided in parentheses and number of participants is shown in brackets.

Table 4: Experiments 2a, 2b, and 2c – Regression Results

	(1) *Forecast Bias* (Experiment 2a)	(2) *Forecast Bias* (Experiment 2b)	(3) *Forecast Bias* (Experiment 2c)	(4) *Forecast Bias* (Pooled)
Left Skew [β_1]	2.09***	1.60***	1.51***	1.90***
	(0.11)	(0.14)	(0.24)	(0.08)
High Salience [β_2]	0.38**	0.03	0.43**	0.29***
	(0.16)	(0.18)	(0.20)	(0.11)
Left Skew x High Salience [β_3]	-0.47**	-0.26	-1.09***	-0.50***
	(0.20)	(0.24)	(0.36)	(0.15)
Constant [β_0]	-0.89***	-0.65***	-0.67***	-0.80***
	(0.08)	(0.10)	(0.15)	(0.06)
# of Forecasts (i.e., Participants)	681	317	114	1,112
R^2	0.41	0.35	0.31	0.38
Remaining Simple Effects				
Left Skew when *High Salience* = 1	1.62***	1.34***	0.41	1.40***
[$\beta_1 + \beta_3$]	(0.17)	(0.20)	(0.27)	(0.12)
High Salience when *Left Skew* =1	-0.10	-0.23	-0.66**	-0.21**
[$\beta_2 + \beta_3$]	(0.12)	(0.17)	(0.29)	(0.10)
Forecast Bias in each Condition				
(1) Right Skew and Low Salience	-0.89***	-0.65***	-0.67***	-0.80***
[β_0]	(0.08)	(0.10)	(0.15)	(0.06)
(2) Left Skew and Low Salience	1.21***	0.94***	0.84***	1.10***
[$\beta_0 + \beta_1$]	(0.07)	(0.10)	(0.19)	(0.05)
(3) Right Skew and High Salience	-0.51***	-0.62***	-0.24*	-0.51***
[$\beta_0 + \beta_2$]	(0.14)	(0.14)	(0.14)	(0.09)
(4) Left Skew and High Salience	1.11***	0.72***	0.18	0.89***
[$\beta_0 + \beta_1 + \beta_2 + \beta_3$]	(0.10)	(0.14)	(0.23)	(0.08)

This table shows the results of regressing *Forecast Bias* – measured relative to the optimal forecast of 6 – on the indicator variables capturing the experimental manipulations and their interaction term for Experiment 2a (column 1), Experiment 2b (column 2), Experiment 2c (column 3), and pooled across experiments (column 4). *Left Skew* equals zero (one) for the conditions with right-skewed (left-skewed) component distributions. *High Salience* equals zero when the aggregate midpoint is not explicitly provided and one when it is explicitly provided (i.e., highly salient).

*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively, referring to two-tailed tests. Standard errors are clustered within participants (see # of Clusters) and presented in parentheses.

APPENDIX A

Experiments 2a, 2b, 2c: Details on Each Wave of Data Collection

Experiment	Population	N*	N	Base	Incentive Pay
2A (Mode)	MTurk	253	191	$2	$8 if forecast correct
2A (Mode)	MBA	116	84	$5	$20 if forecast is correct
2A (Mode)	MBA	95	92	$2	$20 if forecast is correct
2A (Mode)	MBA	104	84	EC	$20 if forecast is correct
2A (Mode)	BBA	308	230	EC	$20 if forecast is correct
2B (Median)	BBA	313	247	-	$10 - $0.50*|forecast - actual|
2B (Median)	MBA	88	70	EC	$10 - $1*|forecast - actual|
2C (Mean)	BBA	186	114	-	$10 - $0.10*(forecast - actual)^2
		1,463	1,112		

N* is the full sample and N is the number of participants who answered all attention checks correctly. EC indicates participants received extra credit for completing the experiment.