



Management Science

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

How Should Time Estimates Be Structured to Increase Customer Satisfaction?

Beidi Hu; , Celia Gaertig, Berkeley J. Dietvorst;

To cite this article:

Beidi Hu; , Celia Gaertig, Berkeley J. Dietvorst; (2024) How Should Time Estimates Be Structured to Increase Customer Satisfaction?. Management Science

Published online in Articles in Advance 20 Dec 2024

. <https://doi.org/10.1287/mnsc.2023.00137>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2024 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

How Should Time Estimates Be Structured to Increase Customer Satisfaction?

Beidi Hu,^{a,*} Celia Gaertig,^b Berkeley J. Dietvorst^a

^aBooth School of Business, University of Chicago, Chicago, Illinois 60637; ^bHaas School of Business, University of California, Berkeley, Berkeley, California 94720

*Corresponding author

Contact: beidi.hu@chicagobooth.edu,  <https://orcid.org/0000-0002-8190-716X> (BH); celia.gaertig@haas.berkeley.edu,  <https://orcid.org/0000-0002-5444-4999> (CG); berkeley.dietvorst@chicagobooth.edu,  <https://orcid.org/0000-0002-8139-8983> (BJD)

Received: April 10, 2023

Revised: February 7, 2024; August 1, 2024

Accepted: August 2, 2024


Published Online in Articles in Advance:
December 20, 2024

<https://doi.org/10.1287/mnsc.2023.00137>

Copyright: © 2024 The Author(s)

Abstract. Businesses across industries, such as food delivery apps and GPS navigation systems, routinely provide customers with time estimates in inherently uncertain contexts. How does the format of these time estimates affect customers' satisfaction? In particular, should companies provide customers with a point estimate representing the best estimate, or should they communicate the inherent uncertainty in outcomes by providing a range estimate? In eight preregistered experiments ($N = 5,323$), participants observed time estimates provided by an app, and we manipulated whether the app presented the time estimates as a point estimate (e.g., "Your food will arrive in 45 minutes.") or a range (e.g., "Your food will arrive in 40–50 minutes."). After participants learned about the app's prediction performance by sampling a set of past outcomes, we measured participants' evaluation of the app. We find that participants judged the app more positively when it provided a range rather than a point estimate. These results held across different domains, different time durations, different underlying outcome distributions, and an incentive-compatible design. We also find that this preference is not simply due to people's dislike of late outcomes, as participants also rated ranges more positively than conservative point estimates corresponding to the upper (i.e., later) bound of the range. These findings suggest that companies can increase customer satisfaction with realized time estimates by communicating the uncertainty inherent in these time estimates.

History: Accepted by Jack Soll, behavioral economics and decision analysis.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Management Science. Copyright © 2024 The Author(s). <https://doi.org/10.1287/mnsc.2023.00137>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Funding: This research was supported by the Bakar Faculty Fellowship at the Haas School of Business at UC Berkeley and the Beatrice Foods Co. Faculty Research Fund at the University of Chicago Booth School of Business.

Supplemental Material: The online appendix and data files are available at <https://doi.org/10.1287/mnsc.2023.00137>.

Keywords: uncertainty • time estimates • time perception • numerical cognition

Introduction

Companies routinely provide customers with time estimates of future events. Everyday examples include Google maps and competitors helping drivers estimate how long trips will take, food delivery apps estimating when food will be delivered, and online shopping websites predicting when goods will arrive. However, the timing of future events is often uncertain. For example, the delivery of restaurant orders and the arrival of Uber drivers can be held up by the randomness inherent in local traffic. And the delivery of goods is subject to flight delays, warehouse problems, and inventory issues. Importantly, many of these factors are outside of the service providers' control. As a result, it is often

impossible to provide customers with precise time estimates that are also perfectly accurate.

Despite the prevalence and significance of this challenge, it is unclear how companies should handle this uncertainty when providing time estimates to customers. Real-life observations suggest that companies currently take different approaches when providing time estimates to customers. For example, some food delivery apps, including Grubhub and Uber Eats, have often shown range estimates to their customers (e.g., "40–50 minutes"), while others, such as DoorDash, have often shown point estimates (e.g., "45 minutes"). In addition, online shopping sites and other platforms sometimes take a hybrid approach, displaying both the most likely

delivery date and a range of possible dates. With some companies presenting point estimates and others presenting range estimates to customers, it is worth considering which leads to greater customer satisfaction: Providing customers with a point estimate, or communicating the uncertainty inherent in the time estimate by providing a range of possible outcomes?

To date, little empirical work has explored how the communication of uncertain time estimates impacts customer satisfaction. And although some research has investigated how consumers react to the communication of uncertain estimates outside of the domain of time, this is likely insufficient to answer the question of how uncertain time estimates will be perceived for several reasons: (a) people may treat time differently from other domains, (b) most of existing work has focused on prospective judgments rather than retrospective judgments after an outcome is revealed, and (c) existing literature does not give a consistent answer regarding people's reaction to uncertain estimates.

In this research, we examine how communicating the uncertainty inherent in time estimates affects customer satisfaction. Because companies can easily change how time estimates are displayed, answering this question will have important practical implications for how companies configure the customer experience.

Point Estimates vs. Range Estimates

Companies need to choose whether and how to communicate the uncertainty inherent in the timing of future events. This often involves a choice between a point estimate and a range estimate. To illustrate, consider the example of a food delivery app. Estimating delivery time requires accounting for the uncertainty in restaurant preparation time, traffic, delays in dispatching, etc. As a result, there are many possible outcomes that could be realized. If all possibilities are represented by a symmetrical distribution centered at 45 minutes, then the food delivery platform could give customers a point estimate (e.g., "45 minutes") or a range (e.g., "35–55 minutes") among other possibilities.¹ The point estimate captures the company's best estimate of what will happen and expresses little uncertainty, while the range communicates that there is uncertainty over multiple future states of the world. How does this decision of whether to communicate uncertainty influence customers' evaluations?

Prior research gives mixed answers to the question of whether companies will benefit from communicating uncertainty to their customers. Most of this research has examined uncertain estimates outside of the domain of time and with a particular focus on *prospective* judgment, that is, judgments occurring *before* outcomes are known.

A large number of these studies suggest that companies may benefit from providing more precise numerical

estimates to their customers. For example, in the advice-giving context, advisees have greater confidence in, attribute more expertise to, and are more likely to choose advisors who provide more precise (and even overprecise) estimates (Price and Stone 2004, Radzevick and Moore 2011, Jerez-Fernandez et al. 2014). People also see companies advertising with precise numbers as more competent (Xie and Kronrod 2012) and consider products labeled with more fine-grained estimates as more trustworthy (Zhang and Schwarz 2012). Companies seem to be aware of this preference for precision and, for example, use precise earnings forecasts as an impression management tactic to signal competence to stakeholders (Hayward and Fitza 2017). Taken together, this evidence suggests that people often reward precise numerical expressions before outcomes are revealed.

However, even before outcomes are known, presenting more precise estimates to customers may not always be advantageous (e.g., Du et al. 2011; Joslyn and LeClerc 2012; Gaertig and Simmons 2018, 2023; Howe et al. 2019; Van Der Bles et al. 2020). For example, in inherently uncertain contexts, people evaluate advisors who provide advice in the form of numerical ranges as no less competent than advisors who provide point estimates (Gaertig and Simmons 2018). Similarly, in the realm of scientific communication, lay audiences consider a scientific claim and its source to be at least as trustworthy when the scientific statements are presented as a numerical range rather than a point estimate (Howe et al. 2019, Van Der Bles et al. 2020). This research suggests that even in contexts where outcomes are not yet realized, people do not always dislike (and sometimes prefer) estimates that acknowledge the uncertainty inherent in future outcomes.

Nevertheless, a crucial distinction between time estimates and many other types of uncertain estimates is that people often experience immediate and transparent outcomes and hence can learn whether these outcomes match their expectations. For example, consumers who order food for delivery naturally learn when their food arrives, and drivers naturally learn how accurate a navigation app's estimate of drive time was. Thus, the present research focuses on *retrospective* judgment in the domain of time, for example, how people evaluate an app *after* learning how the time estimates perform.

The Role of Revealing Outcomes

Communicating the inherent uncertainty in future outcomes may be particularly beneficial in contexts where customers learn about both the estimate and the actual outcome, and hence can judge how well they align with each other. In line with this, research outside the domain of time has found that precise estimates are sometimes penalized when these estimates turn out to be inaccurate (e.g., Tenney et al. 2008, Dieckmann et al. 2010, Radzevick and Moore 2011, Sah et al. 2013,

Pena-Marin and Wu 2019). For example, in a multi-round experiment on financial advice, participants were less likely to choose overprecise (i.e., less calibrated) advisors again in later rounds when they were able to assess the advisors' accuracy (Radzevick and Moore 2011). Similarly, in the context of price estimates, people judge companies that give precise price estimates as less trustworthy than those that give imprecise estimates when these estimates turn out to be incorrect (Pena-Marin and Wu 2019). This work highlights that in contexts in which outcomes are revealed, people may evaluate estimates that express the uncertainty inherent in future outcomes more positively than certain estimates.

Related to these findings, prior work has proposed that when evaluating estimates under uncertainty, people consider both an estimate's *informativeness* and *accuracy* (Yaniv and Foster 1995). *Informativeness* refers to the extent to which the estimate communicates a specific outcome. For example, in the domain of time, a point estimate of 45 minutes is more informative than a range of 35–55 minutes. *Accuracy* refers to the extent to which the estimate aligns with the realized outcome. For example, a point estimate would be considered accurate when the outcome matches the estimate, whereas a range estimate would be considered accurate when the outcome falls within the range. People generally favor estimates that are both informative and accurate (Grice 1975), but these objectives conflict in inherently uncertain contexts and hence consumers face a trade-off (Yaniv and Foster 1995).

In the current work, we use the *accuracy-informativeness* trade-off proposed by Yaniv and Foster (1995) as the starting point for our theorizing, but our empirical investigation moves beyond their framework in several important ways. To illustrate, Yaniv and Foster (1995) tested their framework using uncertain estimates for general knowledge questions. For example, participants in one of their studies saw two different range estimates and indicated which of these estimates they preferred based on the actual answer. Participants were willing to accept some error in accuracy for a more informative (i.e., a narrower) range. In Yaniv and Foster's (1995) studies, however, participants always judged estimates in joint evaluation and the studies did not include any point estimates. Hence, by investigating people's satisfaction with point estimates versus range estimates, we advance our knowledge on how people trade-off accuracy and informativeness when presented with point estimates versus range estimates in separate evaluation. In addition, our empirical investigation focuses on the domain of time, which, to our knowledge, has not been tested. Importantly, people may treat time differently from other domains. In particular, consumers often experience outcomes naturally in the domain of time (e.g., they learn when food was delivered), and hence accuracy judgments may become more important. Indeed, in our

studies, we find that range estimates are favored to point estimates. This holds true as long as the range is not excessively wide, and thus no longer informative.

Along the way, we also provide process evidence that further moves beyond Yaniv and Foster's (1995) framework. In addition to considering the role of accuracy and informativeness, we examine what expectations consumers form about future outcomes when presented with different types of estimates. Specifically, we find that, compared with point estimates, range estimates lead consumers to expect more variability in outcomes, thus making any outcome following a range estimate less likely to violate their expectations.²

Empirical Predictions

We propose that companies choosing whether to present times estimates as point estimates or ranges face a trade-off between *informativeness* and *accuracy* (Yaniv and Foster 1995). Informativeness is desired by customers because it provides a clearer picture of when an event will occur. For example, a customer who learns that "a technician will arrive at your home at 2:30 pm" has a better understanding of what to expect than a customer who learns that "a technician will arrive at your home between 1:30 pm and 3:30 pm" and would prefer the more precise point estimate holding all else constant. However, the timing of events is often inherently uncertain, and as a result, providing more informativeness often comes at the cost of less accuracy. In the example above, given different factors that may influence a technician's arrival time, it is more likely that the technician will arrive between 1:30 p.m. and 3:30 p.m. than precisely at 2:30 p.m. As a result, after the outcome has been realized, the customer could very well be more satisfied having received the less informative but more likely to be accurate range estimate. We posit that in the context of time estimates with realized outcomes, the benefit of increased accuracy will outweigh the cost of less informativeness. More formally:

Hypothesis 1. *When outcomes are revealed, a time estimate presented in the format of a range will lead to greater customer satisfaction than a point estimate.*

Importantly, though, this hypothesis rests on the assumption that the range estimate is still reasonably informative. When a range is extremely wide (e.g., when it covers all possible outcomes), it offers no informational value. Customers are meant to form plans given time estimates, and thus, uninformative time estimates are not useful. For example, a customer receiving a 4-hour window in which dinner could arrive won't know what time to invite guests. At the same time, the marginal improvement in accuracy from making ranges excessively wide is unlikely to counterbalance the decline in informativeness. For

example, if outcomes are normally distributed, providing a range that includes 99% of outcomes instead of 90% of outcomes requires widening the range by approximately 57%³—an immense decrease in informativeness for only a 10% increase in accuracy. Therefore, we predict as a boundary condition that people will not be satisfied with extremely wide ranges. Even though extremely wide ranges will be highly accurate, we believe that people will prefer moderate ranges and even point estimates because they are more informative. Formally:

Hypothesis 2. *A time estimate presented in the format of an excessively wide range will lead to lower customer satisfaction than a reasonably wide range or a point estimate.*

Why would people judge time estimates that express uncertainty more positively than certain estimates? One potential reason, as outlined above, is simply that uncertain estimates are accurate more often than certain estimates because they include more future states of the world. For example, a range (e.g., 40–50 minutes) covers more future states of the world than a point estimate (e.g., 45 minutes), and hence the actual outcome is more likely to fall within the range.

Above and beyond the increase in accuracy, range estimates, by making the uncertainty explicit, may also help set people's expectations so that they are more prepared for variation in future outcomes. That is, customers receiving a range estimate instead of a point estimate may realize that it is uncertain exactly when the outcome will occur and may therefore expect a wider range of outcomes. Since the extent to which expectations are violated matters for people's evaluations of estimates (Teigen and Nikolaisen 2009), people may punish outcomes falling slightly outside of a range less than outcomes differing from a point estimate. In other words, in part because ranges cover more future states of the world, they may expand the interval in which outcomes feel consistent with people's expectations relative to a counterfactual in which a point estimate has been provided. Formally:

Hypothesis 3. *People are more satisfied with range estimates than point estimates after outcomes are revealed because they perceive that range estimates violate their expectations less often.*

The Current Research

In this paper, we report the results of eight experiments investigating whether providing a range or a point estimate leads to greater customer satisfaction with an app. The design of our experiments is inspired by how people interact with digital platforms in the real world, where they experience both the time estimates and the outcomes of the estimates and often see a sequence of individual outcomes. For example, a customer may

place several orders with the same food delivery app over a period of time. To simulate this environment, our studies asked participants to go through multiple trials, in each of which they saw the time estimate and the actual completion time for an app or a model in a specific domain.

In Studies 1–3 and 5–8, participants evaluated time estimates provided by either a food delivery app (Studies 1–3 and 5–7) or a GPS app (Study 8), and we manipulated the type of estimate that participants saw (i.e., point estimate versus range) between subjects. After seeing 20 orders or 20 trips, participants reported their satisfaction with the app. In Study 4, participants evaluated a model estimating the time it would take to complete an incentivized task. Participants in this study went through a series of 10 trials in which they saw one model giving a point estimate and one model giving a range estimate. They were then asked to choose the model they would like to receive advice from for a subsequent incentivized task. In most of our studies, participants always saw the same estimate for all trials, but different outcomes. However, in Study 3, we varied both the time estimates and outcomes across trials, and our results are robust to this design.

Across our studies, we find strong evidence that apps are evaluated more positively when they provide ranges rather than point estimates. Specifically, Studies 1–3 find that ranges are preferred to point estimates, as long as they are not excessively wide (Hypotheses 1 and 2). Study 4 replicates the phenomenon in an incentive-compatible design. Study 5 explores the mechanism of this effect and finds that, compared with point estimates, ranges make outcomes less likely to be perceived as unexpectedly late, and that this perception mediates participants' increased liking of range estimates (Hypothesis 3). In Studies 6 and 7, we extend our investigation to compare ranges with different types of point estimates and find that ranges are even preferred to conservative point estimates that correspond to the upper (i.e., later) bound of a range. Finally, in Study 8, we replicate our findings in a different domain, namely GPS navigation. Throughout the paper, we discuss the results of three additional studies (Studies S1–S3) which complement the main findings.

We report all of our measures, manipulations, and exclusions. Sample sizes were determined before data collection. In all of our studies, we preregistered to exclude all responses from duplicate participant IDs or duplicate IP addresses. In all studies but Study 1, we included a screening attention check at the beginning of the survey and we allowed only those participants who passed the screening check to proceed to the survey. In addition, in all studies but Study 4, we embedded attention check questions at the end of the survey asking participants to indicate key information they saw in the stimuli. These questions are described in

detail in Supplement 1. As preregistered, in the paper, we report the results with all participants, including those who failed these attention check questions. The results excluding those who failed the attention checks do not meaningfully differ, and we report them in Supplement 3. All of our data, materials, and preregistrations are available on ResearchBox: <https://researchbox.org/482>.

Studies 1 and 2: Narrow and Wide Ranges

Studies 1 and 2 test whether point estimates or ranges are preferred in the food delivery domain and whether this depends on the width of the range. We predicted that people would prefer ranges to point estimates, as long as the range is not extremely wide (Hypotheses 1 and 2).

Methods

Participants. We conducted Studies 1 and 2 using U.S. participants from Amazon Mechanical Turk (MTurk). We decided in advance to recruit 700 and 600 participants, respectively. After preregistered exclusions (described above and in Supplement 1), our final samples included 698 and 599 participants, respectively (average age = 39 years,⁴ 44%–50% female).

Procedure. Studies 1 and 2 followed a similar procedure. Participants were asked to evaluate a new food delivery app after seeing both the estimated delivery times and actual delivery times of 20 past orders that people placed with this app. Participants were informed that all 20 orders had the same estimated delivery time, but the actual delivery times varied across orders. That is, we manipulated the estimated delivery times to be

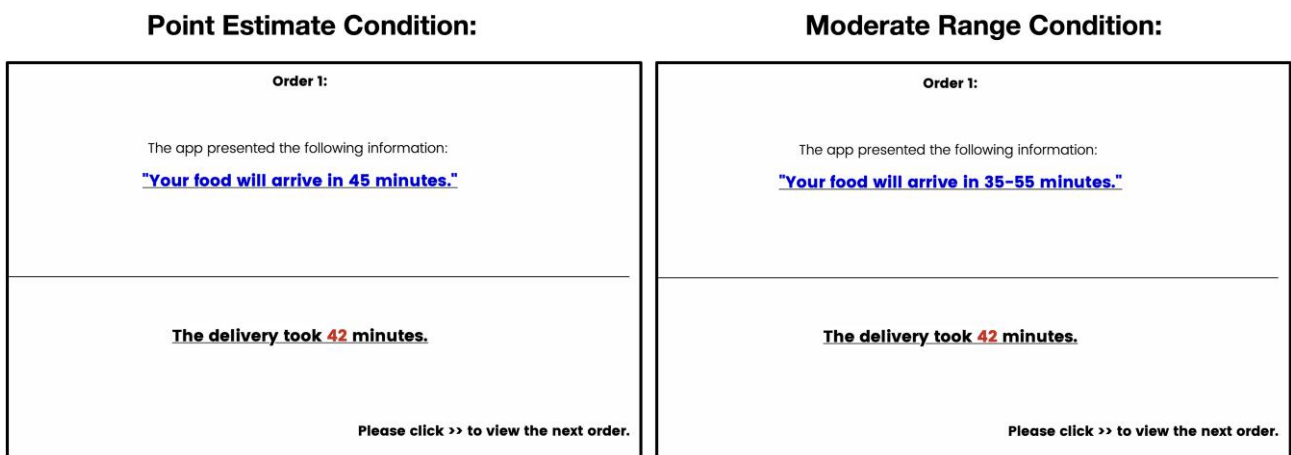
point estimates or ranges between subjects (see manipulations described below), but we kept the estimated delivery time constant across trials. That is, each participant saw the same time estimate for all 20 orders. We made this design choice for Studies 1 and 2 to facilitate participants' learning of the outcome distribution. Notably, in Study 3, we varied both the time estimate and the outcome for individual orders and found results consistent with those of Studies 1 and 2.

After reading the instructions, participants were asked to view the 20 orders sequentially, with each order appearing on the screen one at a time. Participants could click the "continue" button to proceed to the next order only after 2 seconds, ensuring that participants would not skip any orders. Figure 1 shows an example of one order in Study 1 for two of the three conditions included in Study 1.

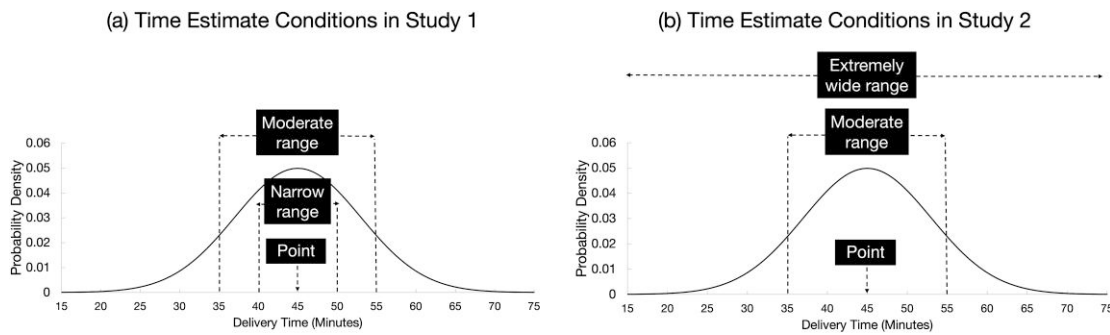
We generated the actual delivery times from a normal distribution with a mean of 45 and a standard deviation of 8. In this distribution, the most likely outcome is 45 minutes. Furthermore, the 80% confidence interval (CI) includes a range of 35 to 55 minutes and the 50% CI includes a range of 40 to 50 minutes, meaning that the actual outcomes fall within these ranges 80% and 50% of the time, respectively. Using this underlying distribution, each participant was randomly assigned to see 20 numbers from this distribution as actual delivery times. For each of Studies 1 and 2, we first generated 50 numbers from the distribution and then randomly assigned each participant to see 20 of these numbers in randomized order to assure that the results are not due to any particular sequence of outcomes.

Manipulations. Figure 2(a) and (b) visualize the conditions included in Studies 1 and 2. Both studies tested whether participants would be more satisfied with the

Figure 1. (Color online) Sample Stimuli Presented in Study 1



Notes. Participants were randomly assigned to see a point estimate ("45 minutes"), a moderate range ("35–55 minutes"), or a narrow range ("40–50 minutes"). The point estimate and the moderate range conditions are pictured above. The actual delivery duration ("42 minutes" in this example) was randomly selected from a normal distribution with a mean of 45 and a standard deviation of 8.

Figure 2. Time Estimate Conditions in Studies 1 and 2

app if it provided ranges rather than point estimates (Hypothesis 1). In Study 1, participants were randomly assigned to see the time estimate presented as a point estimate (“45 minutes”—the mean, median, and mode of the distribution), a moderate range estimate (“35–55 minutes”—the 80% CI), or a narrow range estimate (“40–50 minutes”—the 50% CI).

Study 2 additionally tested whether an extremely wide range (i.e., the 99% CI of the distribution) would be disliked compared with a range of reasonable width and a point estimate (Hypothesis 2). Participants were randomly assigned to see the delivery estimate in the format of a point estimate (“45 minutes”), a moderate range (“35–55 minutes”—the 80% CI), or an extremely wide range (“15–75 minutes”—the 99% CI).

Dependent Measures. After viewing the 20 orders, participants were asked to evaluate the app. Our main measure of interest is participants’ satisfaction with the app. To measure participants’ satisfaction with the app, we asked participants how much they like the delivery estimates provided by the app, how informative⁵ and useful they find them, how accurate they think the app is, and how much they trust the app (all on 7-point scales). We averaged these five items into a single measure of satisfaction with the app ($\alpha \geq 0.92$). Table 1 presents the exact wording of our dependent measures.

In addition, we also measured participants’ perceived accuracy of the app using three different measures that we analyzed separately. Participants indicated what percentage of the app’s estimates they perceived as (1) very accurate and (2) very inaccurate, and (3) by how many minutes, on average, they thought the app was off when it gave an inaccurate estimate. We presented the satisfaction scale and perceived accuracy measures to participants in counterbalanced order.

At the end of the survey, participants reported their age and gender.

Results and Discussion

Analysis Plan. In both studies, we preregistered that we would conduct any tests comparing participants’

responses in individual conditions using ordinary least squares (OLS) regressions with dummies indicating the condition. For example, in Study 1, we first regressed the satisfaction measure on dummies indicating (1) the narrow range condition and (2) the moderate range condition to compare these two conditions to the point estimate condition. Second, we regressed the satisfaction measure on dummies indicating (1) the point estimate condition and (2) the narrow range condition to compare these conditions to the moderate range condition.

We present the results of the satisfaction measure below, which is our main measure of interest. We present the results for individual measures of perceived accuracy for Studies 1 and 2 (and also for Studies 3 and 6–8) in Supplement 2 and discuss them in the General Discussion.

Satisfaction with the App. Table 2 shows the results for the satisfaction measure. Providing evidence for Hypothesis 1, in Study 1, we found that participants were more satisfied with the app both when the delivery estimates were presented as moderate ranges (“35–55 minutes”) rather than point estimates (“45 minutes”), $b = 0.56$, $SE = 0.10$, $p < 0.001$, and when they were presented as narrow ranges (“40–50 minutes”) rather than

Table 1. Satisfaction with the App Measured in Studies 1–3 and 5–8

Satisfaction with the App ($\alpha \geq 0.92$)

1. How much do you like the delivery (arrival) estimates provided by this app? (1 = Not at all, 7 = Extremely)
2. How informative do you find the delivery (arrival) estimates provided by this app? (1 = Not at all informative, 7 = Extremely informative)
3. How useful do you find the delivery (arrival) estimates provided by this app? (1 = Not at all useful, 7 = Extremely useful)
4. How accurate do you think this app is? (1 = Not at all accurate, 7 = Extremely accurate)
5. How much do you trust this app? (1 = Not at all, 7 = Extremely)

Note. Text in parentheses was displayed for the GPS app evaluation in Study 8.

Table 2. Satisfaction Results of Studies 1 and 2

	Point estimate (45 minutes)		Moderate range (35–55 minutes)		Narrow range (40–50 minutes)		Wide range (15–75 minutes)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Study 1	4.81	1.28	5.37	1.08	5.32	1.00	—	—
Study 2	4.62	1.41	5.31	1.11	—	—	4.32	1.50

point estimates, $b = 0.50$, $SE = 0.10$, $p < 0.001$. The two range conditions did not significantly differ from each other, $b = 0.06$, $SE = 0.10$, $p = 0.571$.

In Study 2, we replicated Study 1’s results that participants liked the app more when they saw moderate ranges (“35–55 minutes”) rather than point estimates (“45 minutes”), $b = 0.68$, $SE = 0.13$, $p < 0.001$. In addition, and critical to Hypothesis 2, participants liked the app less when it presented an extremely wide range (“15–75 minutes”) rather than a moderate range ($b = -0.99$, $SE = 0.13$, $p < 0.001$) and even a point estimate ($b = -0.30$, $SE = 0.14$, $p = 0.026$). This is not surprising, as the extremely wide range can be considered as being no longer informative of when the food would arrive.

The results from Studies 1 and 2 suggest that choosing between a point estimate and a range estimate involves trading off accuracy and informativeness (Yaniv and Foster 1995). Point estimates communicate a specific outcome and are easy to understand, but they may convey a false sense of certainty regarding the predicted time and are more likely to be wrong. Ranges, on the other hand, are more likely to contain the actual outcome, but they are also less informative because they indicate a less specific outcome. Our results suggest that people seem to prefer the increased accuracy of ranges compared with point estimates under the conditions that we tested (Hypothesis 1). This seems to hold true as long as ranges are not excessively wide, and hence no longer informative (Hypothesis 2).

Combining Point and Range Estimate

In Studies 1 and 2, we tested whether people prefer an app to display point estimates or range estimates. But companies may also choose to display a combination of these types of estimates to their customers. We conducted another study ($N = 578$; see Study S1 in Supplement 5) to investigate how the combined display of a point estimate and a range estimate influences customer satisfaction. We presented participants in this study with a screen mockup of a food delivery app that closely resembled app interfaces that customers might encounter in the real world (see Supplement 5). We manipulated whether the app displayed a point estimate (“45 min”), a range estimate (“35–55 min”), or a combination of both (“45 min (35–55 min)”).

Replicating the results from Studies 1 and 2, participants were more satisfied with the app when it

presented a range estimate than when it presented a point estimate ($p < 0.001$). In addition, participants were also more satisfied with the app when it presented both a point estimate and a range estimate than when it presented only a point estimate ($p < 0.001$). However, participants did not rate the app as more positively when it presented both a point estimate and a range estimate ($M = 4.97$, $SD = 1.23$) than when it presented only a range estimate ($M = 5.12$, $SD = 1.14$), $p = 0.253$. Moreover, participants perceived the combined display of a point estimate and a range estimate as *less* accurate than the range estimate alone ($ps \leq 0.038$, see Supplement 5 for details). These results suggest that even though displaying a point estimate along with a range estimate affords both informativeness and accuracy, this configuration does not increase customer satisfaction compared with presenting a range estimate alone. On the contrary, adding a point estimate to a range estimate did not affect ratings of customer satisfaction and even reduced the estimate’s perceived accuracy. These findings provide further support for our theorizing that in the domain of time where people naturally experience outcomes, they may especially value the increased accuracy that range estimates offer and benefit from estimates that increase their expectation of uncertainty.

Study 3: Different Underlying Distributions

In Studies 1 and 2, participants saw the same delivery estimates across all 20 orders, and the actual delivery times were drawn from the same underlying distribution. In the real world, however, customers may experience food delivery orders with different outcome distributions, and consequently, different delivery estimates. We conducted Study 3 to test the robustness of our prior findings to a design in which the 20 orders have different delivery estimates drawn from different distributions.

Methods

Participants. We conducted Study 3 using U.S. participants from Prolific Academic. We decided in advance to recruit 500 participants. After preregistered exclusions, our final sample included 479 participants (average age = 41 years, 49% female).

Procedure. Study 3 followed a similar procedure as Studies 1 and 2 except for two changes. First, participants

were only assigned to one of two conditions: point estimates versus range estimates, with the point estimates corresponding to the mean and the range estimates corresponding to the 80% CI of the underlying distribution. Second, instead of showing participants the same time estimate from the same distribution for all 20 orders, participants saw a different time estimate from a different distribution for each of the 20 orders. Specifically, participants saw an estimate and an outcome from 20 unique outcome distributions (presented in random order), with means ranging from 36 to 55 minutes, each with a standard deviation (SD) of 8 (and hence an 80% CI encompassing 20 minutes). Participants assigned to the point estimate condition saw a point estimate corresponding to the mean of the distribution (e.g., “Your food will arrive in 39 minutes” for the distribution with a mean of 39), and participants assigned to the range estimate condition saw an 80% confidence interval centered at the mean of the distribution (e.g., “Your food will arrive in 29–49 minutes” for the distribution with a mean of 39). For each trial, the actual delivery outcome was randomly drawn from the corresponding underlying distribution in the same way as in Studies 1 and 2.

Results and Discussion

Satisfaction with the App. As preregistered, we regressed the satisfaction measure on the range estimate condition (contrast-coded; $+0.5 =$ range estimate, $-0.5 =$ point estimate). Participants liked the app more when they saw range estimates ($M = 5.32$, $SD = 1.12$) than point estimates ($M = 4.70$, $SD = 1.26$) across the 20 orders, $b = 0.61$, $SE = 0.11$, $p < 0.001$. Study 3 thus replicates the findings in Studies 1 and 2 with a design closer to real-world contexts where the same person may encounter different delivery estimates from the same app on any given day.

Study 4: Incentive-compatible Decision

Studies 1–3 find that after receiving information on how time estimates perform, people are more satisfied with a food delivery app that presents time estimates as ranges rather than point estimates. It follows that, when given a choice, people should also be more likely to choose to receive time estimates as ranges rather than point estimates, particularly if they had a chance to see the time estimates perform. Study 4 tests this prediction in an incentive-compatible design. Participants chose between receiving time estimates from one of two models, with one providing point estimates and the other providing range estimates. Participants could use the model’s time estimates to decide on the allotted time for an incentivized effort task later in the survey. Successful completion of the effort task within the allotted time determined participants’ bonus payments. Notably, the incentives were designed to reflect the

trade-off between allocating excessive time (resulting in wasted time) versus allocating insufficient time (resulting in incomplete tasks), a challenge people frequently encounter when estimating time in the real world. We predicted that participants would be more likely to choose the model offering range estimates.

Methods

Participants. We conducted Study 4 using U.S. participants from Prolific Academic. We decided in advance to recruit 400 participants. Because of the elaborate design of this study, we decided to start the survey with an instructions phase that included four comprehension check questions which we describe in the *Comprehension checks* section below (see also Supplement 1). As preregistered, participants who failed any comprehension check more than once ($N = 47$) were automatically excluded and did not complete the rest of the study. A total of 399 participants completed the study. We preregistered to exclude responses from IP addresses or Prolific IDs that appeared more than once in our data set (7 exclusions). Our final sample included 392 participants (average age = 44 years, 50% female).

Design. Participants in this study were asked to decide how much time they would like to allot for themselves for an effort task that consisted of moving sliders to a specified number. We designed the specifics of this task such that participants were incentivized to set the time periods as accurately as possible (as explained in the *Procedure* section below). To help participants determine how much time they might need for this task, we introduced two models, Model Alpha and Model Beta, each of which gave an estimate for the completion time of this task based on past participants’ data. The models differed in the format in which they provided the time estimate, with one model providing point estimates and the other model providing range estimates. Participants first gained experience with both models by seeing the models’ estimates for one variant of the task and the actual completion time of ten past participants. They then decided which model to receive advice from when setting their own completion time for the slider task.

Procedure. At the beginning of the survey, participants first learned that they would complete three pages of slider tasks later in the study. Each page would contain 5–25 sliders labeled with numeric values, and they would need to move the sliders to the values indicated. Participants saw a visual demonstration of the task (Figure 3). Notably, we designed the task such that its completion requires no specific skills other than pure effort, which also means that past participants’ completion times are informative of a future participant’s completion time.

Figure 3. (Color online) A Visual Demonstration of the Sliders Task Presented in the Study 4 Instructions



Notes. Participants read that “For each slider, you will be asked to use the mouse to move the slider to the value indicated. That is, you should move the slider until the value shown on the handle matches the target value on the left, like the first slider in the screenshot below.”

Participants read that before they start the slider task, they would first decide how much time to allocate to themselves for completing the task, and that the time they decide to allocate would determine the exact amount of time they would have for each page. Participants learned that if they correctly placed all sliders on a randomly selected page within the allotted time period, they would receive an additional \$1 bonus.

Participants further learned that (a) the slider page would auto-advance after the allotted time regardless of their progress on the tasks, and (b) they cannot advance to the next page until the allotted time had elapsed, even if they finished the task. We made these specific design choices to incentivize participants to set the allotted time to be just above their actual completion time. In fact, we made clear that the implication of this is that “you should give yourself enough time to finish the task, but not too much time or you will have to waste time waiting.” Our goal in designing these incentives was to capture the trade-off that people often make when generating time estimates in real life—allocating too much time and waiting versus not allocating enough time and failing to complete a goal. For example, in the context of driving, people who allocate too much time to travel to an appointment may have to waste time waiting once they have arrived, and people who allocate too little time may arrive late to their appointment. Finally, in the instructions, participants also learned that once the page auto-advances, the timer for the next page would automatically start a few seconds later, and so they must complete the three pages consecutively with only a few seconds in

between. This design choice was aimed at preventing online participants from switching tabs and navigating away during the study. We asked participants comprehension check questions to ensure that they understood all of this, and we describe them in the *Comprehension checks* section below.

To help participants generate their own time estimates, we introduced two models, Model Alpha and Model Beta, both of which used data from past participants to predict the completion time for each page of sliders that participants were asked to complete. Participants learned that they would first see how the models performed previously, meaning that they would see the time estimates that each model provided and how long it actually took participants to complete a certain number of sliders. One model gave point estimates, the other model gave range estimates, and we counterbalanced whether Model Alpha or Beta gave point estimates or range estimates.

To obtain the model estimates for this training stage of the survey and for the main task, we ran a pretest ($N = 220$ on Prolific) where participants completed several pages of the slider task, each with a different number of sliders. We incentivized the pretest participants to be as accurate as possible to ensure that they were exerting a similar amount of effort (and time) in the task as participants in the main study. Table 3 shows the point estimates and range estimates for the number of sliders that we used in the main study based on this pretest. Because for each number of sliders the distribution of the pretest data was right-skewed (i.e., the bulk of the completion time fell on the lower end of the

Table 3. Model Estimates Presented in Study 4

		Number of sliders	Point estimate	Range estimate
Training stage		10 Sliders	54 seconds	45–71 seconds
Incentivized stage	Slider page 1	8 Sliders	46 seconds	38–66 seconds
	Slider page 2	12 Sliders	66 seconds	54–79 seconds
	Slider page 3	18 Sliders	90 seconds	77–114 seconds

Notes. 1. We used the median of the duration data from a pretest ($N = 220$) as the point estimate, and the 25th percentile to 75th percentile (i.e., 50% CI) as the range estimate. 2. While participants read that the number of sliders on each page may range from 5 to 25, we presented all of them (after our DV) with three pages of sliders with 8, 12, and 18 sliders, respectively.

distribution, with a few participants taking exceptionally longer to finish), we took the median of the distribution to be the point estimate and the 25th to 75th percentile (i.e., 50% CI) to be the range estimate.⁶

Participants gained experience with how the models had performed previously by observing the models' estimates for moving ten sliders and the actual completion time from ten randomly selected past participants from the pretest. Based on the pretest, the models' estimates for moving ten sliders were 54 seconds and 45–71 seconds in the point estimate and range estimate conditions, respectively (see Table 3), though the actual completion times of the ten randomly selected past participants deviated from these estimates. Participants saw the information for each of the ten past participants on a separate page, one after the other. That is, on each page, participants saw Model Alpha's estimate for moving ten sliders, Model Beta's estimate, and how long it took a randomly selected past participant to actually complete the ten sliders. Figure 4 shows an example of one such page. This stage of the study was similar to Studies 1–3 in which participants gained experience with the models' time estimates before indicating their preference for a time estimate.

After observing the performance of both models, participants were reminded that they can select one of the models to give them estimates for the slider task they were about to complete, and that they would receive a \$1 bonus payment if they correctly placed all sliders within the allotted time on one randomly selected page. Participants saw each model's estimates for moving ten sliders again (54 seconds versus 45–71 seconds) and were asked, "For the next 3 pages of slider tasks you will complete, which model's advice would you like to receive?" (Options: "Model Alpha" "Model Beta"). This choice was our main dependent measure.

After participants made their choices, the rest of the study proceeded as what we described: Participants

saw their chosen model's estimate for each of three pages of sliders and were asked to set their own estimate for each page. Note that it was only at this point that they learned that the number of sliders they had to complete on pages 1–3 was 8, 12, and 18, respectively, and they saw their chosen model's estimate for these numbers of sliders (see Table 3 for the model estimates). For example, participants who chose the range estimate model read that for page 1 they would be moving 8 sliders, and that the model gave the estimate of "38–66 seconds." In contrast, participants who chose the point estimate model read that the model gave the estimate of "46 seconds." On the same page where they saw the model's estimate, participants indicated how much time they would like to allot for themselves for this page of the slider task by entering a duration (in seconds) into a text box. Participants decided on their allotted time for all three pages of sliders and then proceeded to start with the first page of slider task.

While participants were completing each page of sliders, a countdown timer set to the allotted time stayed at the top of the page and showed the remaining time left on this page. Once the timer hit zero, the survey automatically advanced and participants saw a transition page informing them that their next page would start in three seconds. After three seconds, participants were automatically directed to the second page of slider task with the timer counting down from their allotted time (see Supplement 4 for screenshots from the survey). After completing all three pages of slider task, participants reported their age and gender at the end of the survey.

Comprehension Checks. At the beginning of the survey and before participants started the task, we tested whether they understood the critical elements of our study design with two sets of two comprehension checks, one set after participants learnt about the slider

Figure 4. Sample Stimulus Presented in Study 4

Participant 1:

Model Alpha's Estimate (for moving 10 sliders): 54 seconds

Model Beta's Estimate (for moving 10 sliders): 45–71 seconds

Actual Completion Time: 50 seconds

Please click >> to view the next participant.

Notes. On each page, participants saw both models' estimates for moving ten sliders and the actual completion time by one past participant. They saw the information of ten participants in total. We counterbalanced which model gave a point estimate or a range estimate. The actual completion time was randomly selected from participants' duration data in the pretest.

task and the incentive, and the other after they learnt about the choice between two models. Specifically, these questions asked participants whether they understood (1) that the allotted time would be the exact amount of time they spend on each page, (2) that once they finish one page, the next slider page would start automatically after a few seconds, (3) that their model choice applies to all three pages of sliders that they are asked to complete and that they would determine the allotted time for all pages before starting with their first page, and (4) that they would earn a bonus if they correctly placed all sliders on a randomly selected page. Supplement 1 presents the exact wording of these questions. Participants were given two attempts to pass each of these comprehension check questions, and 98.0%, 98.4%, 93.5%, and 99.3% of participants passed questions 1, 2, 3, and 4 within two attempts, respectively. Participants who failed any comprehension check more than once ($N = 47$) were automatically excluded and did not complete the rest of the study.

Results and Discussion

Model Choice. We preregistered that we would code participants' choice of the model as 1 if they chose Model Alpha and 0 if they chose Model Beta. We used OLS to regress participants' choice of Model Alpha on whether Model Alpha gave a range estimate or a point estimate (+0.5 = range estimate, -0.5 = point estimate). Participants chose Model Alpha significantly more often when it gave range estimates (74.6%) instead of point estimates (33.3%), $b = 0.41$, $SE = 0.05$, $t = 8.99$, $p < 0.001$. We analyzed the model choice with OLS because the coefficient is easier to interpret (i.e., as the percentage point difference between conditions); a logistic regression yielded similar results, $OR = 5.88$, $SE = 1.31$, $z = 7.93$, $p < 0.001$.

For ease of interpretation, we also converted the dependent measure to participants' choice of the model giving range estimates. This analysis shows that 70.7% of participants chose the model giving range estimates over the one giving point estimates, and this choice proportion was significantly different from 50%, $p < 0.001$ in a two-sided binomial test. These results suggest that after gaining experience with both types of estimates, people prefer to receive time estimates as ranges rather than point estimates. This implies that, all else equal, companies may have a better chance of satisfying customers in the long term if they provide time estimates as ranges rather than point estimates.

Exploratory Analyses. We conducted additional exploratory analyses to examine whether participants' choice of point versus range estimates affected the amount of time they allocated for themselves in the slider task and ultimately their chance of receiving the bonus payment.

Participants allotted significantly longer time periods if they chose the model giving range estimates (Median: 60, 70, and 100 seconds for the pages of 8, 12, and 18 sliders, respectively) than if they chose the model giving point estimates (Median: 50, 66, and 90 seconds for the pages of 8, 12, and 18 sliders, respectively), $t_s > 5.08$, $p_s < 0.001$ for each of the three pages. The accuracy results were consistent with this: On average, those choosing the range estimates model submitted 2.08 completely correct pages (and were 69.3% likely to receive the bonus payment), which was directionally higher than those choosing the point estimates model (on average 1.86 correct pages, 62.0% likely to receive the bonus payment), $t(390) = 1.79$, $p = 0.075$. All told, these exploratory results suggest that participants tend to set more conservative time estimates after receiving advice from the range estimates model and that directionally increased their chance of receiving the bonus.

Overall, Study 4 finds that after getting feedback on both types of time estimates, participants preferred receiving advice from a model giving range estimates rather than from a model giving point estimates for a subsequent incentivized task. Thus, Study 4 conceptually replicates the results of Studies 1–3 with an incentive-compatible design and provides additional support for the notion that people prefer range estimates to point estimates in the domain of time.

Study 5: The Role of Perceived Violation of Expectations

Studies 1–4 find evidence that people prefer time estimates as ranges rather than point estimates after outcomes are revealed. Study 5 explores the potential mechanism for this preference. We propose that the format of the estimate (range versus point estimate) is likely to not only affect customers' expectations about when an event will take place but may also influence their expectations about the uncertainty in future outcomes. This may make the same outcome more or less likely to violate expectations depending on the format of the initial estimate. A point estimate specifies one specific point in time, conveying little variance, and hence any outcome noticeably deviating from this estimate may violate people's expectations, for example, leading people to perceive an outcome that exceeds the point estimate as unexpectedly late. In contrast, a range estimate widens the range of expected outcomes and explicitly expresses the uncertainty in the timing of future outcomes, which may make any outcome less likely to violate people's expectations (Hypothesis 3).

To examine the antecedent of this proposition—that people anticipate a wider range of outcomes when they see a range estimate compared with a point estimate—we conducted Study S2 ($N = 389$; see Supplement 6 for details). In Study S2, we provided participants with a

food delivery estimate either as a point estimate (43 minutes) or a range estimate (33–53 minutes). Participants then constructed their expected outcome distributions by indicating how likely they thought the actual delivery time would be to fall into different time intervals using a distribution builder tool (André 2016, Hu and Simmons 2024). We present the results in Figure 5. Participants constructed distributions with significantly larger standard deviations when they saw a range estimate than when they saw a point estimate ($p < 0.001$). In other words, in prospective evaluations (i.e., before outcomes are realized), range estimates seem to expand people's expected range of outcomes compared with point estimates.

Consequently, after outcomes are realized, we expect people to perceive various outcomes as less likely to violate their expectations when initially given a range estimate than a point estimate. We test this proposal in Study 5. Specifically, we examine whether food deliveries are less likely to be perceived as unexpectedly late when the app provides a range rather than a point estimate. We further test whether people's perceptions of the outcome being unexpectedly late mediate the difference in their satisfaction with ranges and point estimates. Importantly, in the food delivery domain, early arrivals are often innocuous, but late arrivals may be associated with higher costs (e.g., staying hungry). Thus, we expect people to be particularly dissatisfied with unexpectedly late orders, and we posit that the perceived prevalence of such unexpectedly late orders will predict customer satisfaction with this app.

Methods

Participants. We conducted Study 5 using U.S. participants from Prolific Academic. We decided in advance to recruit 500 participants. After our preregistered

exclusions, our final samples included 468 participants (average age = 37 years; 49% female).

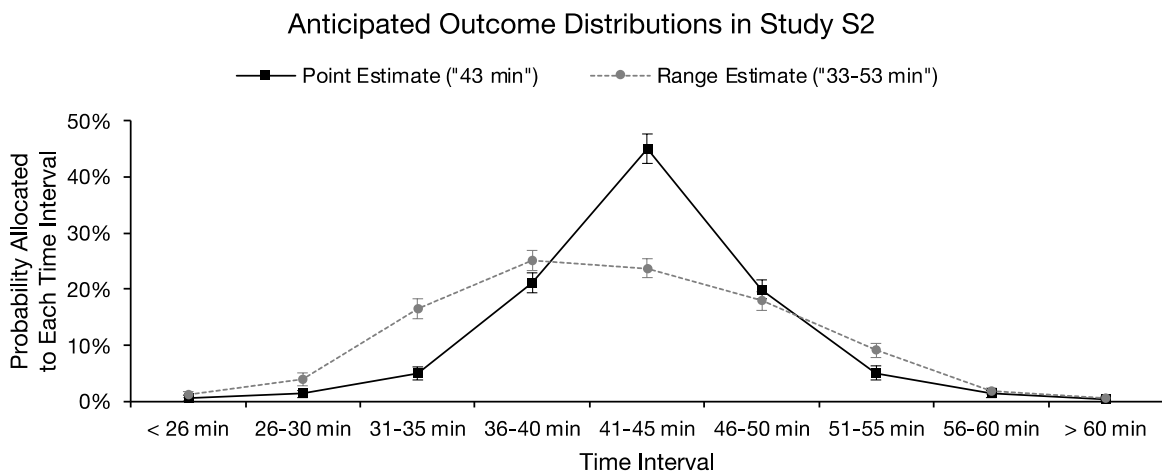
Procedure. The procedure for Study 5 was the same as in Studies 1 and 2 except for three changes. First, participants were randomly assigned to one of two conditions: point estimate (45 minutes) versus range (35–55 minutes). Second, participants in this study only reported their satisfaction with the app by responding to the items in the satisfaction measure (see Table 1), and they were not asked to complete additional measures related to their perceived accuracy of the app. Third, we included an additional question that was aimed at assessing the extent to which outcomes are perceived as being unexpectedly late. Participants were asked, "How often are deliveries on this app unexpectedly late?" (1 = Almost never, 7 = Almost always). We hypothesized that this measure would mediate the effect of estimate format on condition. We counterbalanced the order in which the mediator question and the satisfaction questions were presented.

Results and Discussion

Satisfaction with the App. As preregistered, we regressed the satisfaction measure on the range estimate condition (contrast-coded; +0.5 = range estimate, -0.5 = point estimate). Replicating the results from Studies 1–3, participants again expressed greater satisfaction with ranges ($M = 5.16$, $SD = 1.06$) than point estimates ($M = 4.34$, $SD = 1.28$), $b = 0.82$, $SE = 0.11$, $p < 0.001$.

Mediation Analysis. We next used bootstrapped mediation to examine whether people's perceptions of the outcomes being unexpectedly late mediate the effect of condition on satisfaction. Participants judged the outcomes to be unexpectedly late less often when they

Figure 5. Study S2 Results



Notes. Average distributions of participants' anticipated delivery time given a point estimate (43 minutes) or a range estimate (33–53 minutes). Error bars represent 95% confidence intervals. (See Supplement 6 for details.)

saw range estimates ($M = 2.47$, $SD = 1.18$) rather than point estimates ($M = 4.14$, $SD = 1.10$), $b = -1.67$, $SE = 0.11$, $p < 0.001$. Moreover, participants who judged outcomes to be unexpectedly late less often were more satisfied with the app, $b = -0.44$, $SE = 0.04$, $p < 0.001$. After controlling for participants' judgments about how often outcomes were unexpectedly late, the effect of condition on overall satisfaction was no longer significant (path c' : $b = 0.13$, $SE = 0.12$, $p = 0.286$; path c : $b = 0.82$, $SE = 0.11$, $p < 0.001$). A bootstrapped mediation analysis revealed a significant indirect effect of the range versus point estimate on overall satisfaction via the mediator: indirect effect = 0.69, 95% CI = [0.51, 0.87]. This result is consistent with the notion that people's preference for ranges over point estimates relates to ranges violating their expectations less frequently.

Additional Evidence. We conducted another study (Study S3, $N = 188$; see Supplement 6 for details) that supports the notion that people view more outcomes as consistent with a range estimate than a point estimate. In this study, we again provided participants with either a point estimate (43 minutes) or a range estimate (33–53 minutes). But instead of reporting the extent to which their expectations were violated, participants in this study indicated, for a series of possible outcomes, whether they would consider the outcome to be consistent with the estimate provided to them. In line with the results of Study 5, participants identified more outcomes as consistent with the range estimate than the point estimate ($p < 0.001$). For example, an outcome of “26–30 minutes,” which falls outside of the range of 33–53 minutes, was rated as consistent with the range estimate 23.96% of the time, but as consistent with the point estimate only 11.96% of the time. We provide the full results of this study in Supplement 6.

Taken together, the results of Study 5 and Studies S2 and S3 suggest that a range estimate leads people to expect more variability in outcomes than a point estimate. In line with this, in Study 5, we find that after seeing a series of outcomes following a range estimate,

people are less likely to feel that their expectations have been negatively violated compared with those who saw a point estimate. Our mechanism evidence suggests that people's increased satisfaction with ranges versus point estimates may be related to the lower frequency with which a range estimate negatively violates their expectations.

Studies 6 and 7: Optimistic and Conservative Estimates

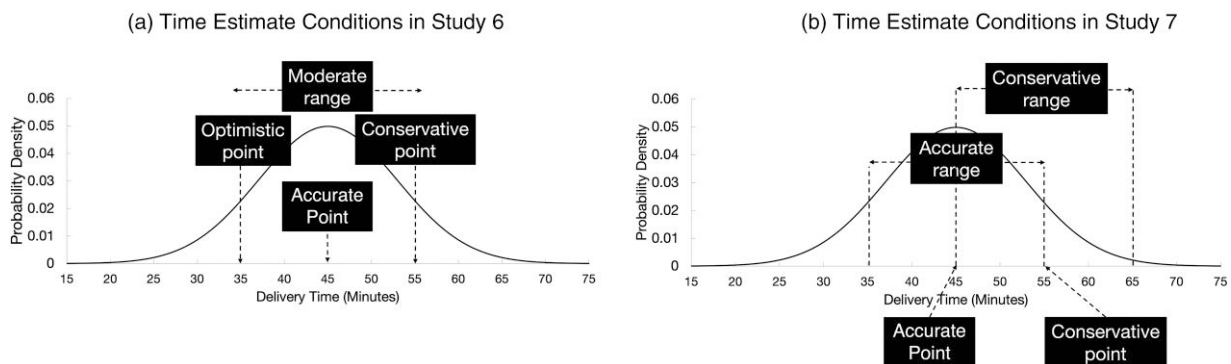
In Studies 6 and 7, we extend our investigation to the question of where to place estimates in the distribution of potential outcomes by manipulating the estimates' location in the underlying distribution. This also allows us to test whether people are as satisfied with a conservative point estimate that corresponds to the upper (i.e., later) boundary of the range as they are with the range itself. In other words, we test an alternative account that people like ranges because they compare outcomes to the upper boundary of the range and only penalize (objectively) late arrivals that cross this boundary. If people only focus on the upper bound, then they should like a conservative point estimate that matches the upper bound of the range just as much as the range.

Methods

Participants. We conducted Studies 6 and 7 using U.S. participants from MTurk. We decided in advance to recruit 800 and 1,000 participants, respectively. After our preregistered exclusions, our final samples included 801 and 998 participants, respectively (average age = 39 years, 44%–52% female).

Procedure. Studies 6 and 7 followed the same procedure as Studies 1 and 2 except that we included different conditions that we describe in detail below. Figure 6(a) and (b) illustrate the manipulations used in Studies 6 and 7. In addition to further testing Hypothesis 1 that ranges are preferred to point estimates, Studies 6 and 7 investigate how estimates at different locations in the

Figure 6. Time Estimate Conditions in Studies 6 and 7



distribution of potential outcomes influence customer satisfaction.

In Study 6, we examined participants’ preferences for point estimates located at different points in the distribution. In addition to including an accurate point estimate (“45 minutes”—mean, median, and mode of the distribution) and a moderate range estimate (“35–55 minutes”), we also included an optimistic point estimate (“35 minutes”) and a conservative point estimate (“55 minutes”). The optimistic and conservative point estimates corresponded to the lower bound (10th percentile) and upper bound (90th percentile) of the range, respectively. Participants were randomly assigned to one of the four estimate conditions (see also Figure 6(a)).

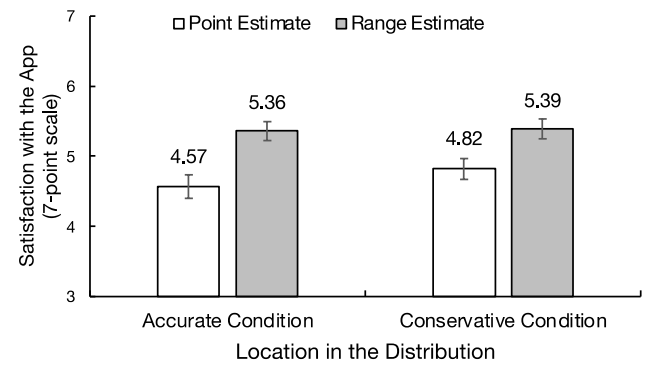
In Study 7, we orthogonally manipulated whether participants saw a point estimate or a range estimate, and whether that estimate was accurate or conservative. That is, we randomly assigned participants to one of the four conditions in a 2 (format: point estimate versus range estimate) × 2 (location in the distribution: accurate versus conservative) between-subjects design. Participants either saw an accurate point estimate (“45 minutes”), an accurate range estimate (“35–55 minutes”), or they saw a conservative point estimate (“55 minutes”) corresponding to the upper bound of the range, or a conservative range estimate (“45–65 minutes”) centered at the conservative point estimate (see also Figure 6(b)).

Results and Discussion

Analysis Plan. We analyzed the results of Study 6 using the same regression approach as in Studies 1 and 2. That is, we regressed the dependent measure on condition dummies comparing participants’ responses in individual conditions. In Study 7, we regressed the dependent measure on (1) the point estimate condition (contrast-coded), (2) the accurate condition (contrast-coded), and (3) their interaction.⁷

Study 6. Table 4 shows the results for Study 6. Participants in Study 6 again liked ranges (“35–55 minutes”) more than accurate point estimates (“45 minutes”), $b = 0.87, SE = 0.14, p < 0.001$. Importantly, however, participants also rated ranges more positively than conservative point estimates corresponding to the upper bound of the range (“55 minutes”), $b = 0.67, SE = 0.14, p < 0.001$. This suggests that the preference for ranges is not simply due to people disliking deliveries that are

Figure 7. Results of Study 7



Note. Error bars represent 95% confidence intervals.

objectively late (i.e., deliveries falling beyond the provided estimates) as the conservative point estimate and the range estimate lead to objectively late outcomes with the same frequency.

Among the point estimates at different locations, participants were less satisfied with the optimistic point estimates (“35 minutes”) than both the accurate ($b = -1.47, SE = 0.14, p < 0.001$) and conservative point estimates ($b = -1.68, SE = 0.14, p < 0.001$). Furthermore, they liked the conservative point estimates no less than the accurate point estimates, $b = 0.20, SE = 0.14, p = 0.164$. Put differently, although people don’t seem to like a food delivery app that constantly underestimates delivery times, they seem to not dislike it when the app overestimates delivery times. It is worth noting that these results diverge from past theories (e.g., Yaniv and Foster 1995) which predict that accurate point estimates should always be preferred over conservative point estimates that are equally informative but objectively less accurate.

Study 7. Figure 7 shows the results for Study 7. Participants in Study 7 evaluated ranges more positively than point estimates ($b = 0.68, SE = 0.08, p < 0.001$), both when the estimates were centered at the middle of the distribution (accurate condition; $b = 0.79, SE = 0.11, p < 0.001$) and when they fell on the upper tail of the distribution (conservative condition; $b = 0.57, SE = 0.11, p < 0.001$). There was no significant main effect of location ($p = 0.074$) nor a significant interaction between estimate type and location ($p = 0.161$).

Although there was no main effect of location, Figure 7 shows that in this study, participants preferred conservative (“55 minutes”) to accurate *point* estimates

Table 4. Satisfaction Results of Study 6

Point estimate (45 minutes)		Conservative point estimate (55 minutes)		Optimistic point estimate (35 minutes)		Range (35–55 minutes)	
<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
4.53	1.51	4.73	1.41	3.05	1.66	5.40	1.10

“45 minutes”), $b = 0.25$, $SE = 0.12$, $p = 0.037$. In addition, and in line with Study 6 results, we again found that participants rated accurate ranges more positively than conservative point estimates ($b = 0.54$, $SE = 0.11$, $p < 0.001$), suggesting that people’s preference for ranges is not simply due to an aversion to deliveries falling beyond the estimates.

Taken together, Studies 6 and 7 find that ranges lead to greater customer satisfaction than point estimates, regardless of where in the distribution the point estimates are located, further supporting Hypothesis 1. Importantly, because outcomes fell beyond conservative point estimates (55 minutes) just as often as they fell beyond accurate ranges (35–55 minutes), participants’ preference for ranges cannot be simply explained by an aversion to objectively late outcomes.

Study 8: A Different Domain and Different Time Durations

In Study 8, we extend our investigation to a different context (driving trip duration estimates) and different time durations, and we also use a different underlying outcome distribution. This allows us to test whether the results we have found so far generalize to a different domain and different time durations.

Method

Participants. We conducted Study 8 using U.S. participants from Prolific Academic. We decided in advance to recruit 900 participants. After our preregistered exclusions, our final sample included 888 participants (average age = 39 years; 50% female).

Procedure. The procedure for Study 8 was very similar to that of Studies 1–2 and 6–7, with three notable exceptions. First, to examine whether our results generalize to a new domain, participants evaluated a GPS app instead of a food delivery app. They were presented with both the estimates and the actual arrival times of 20 past trips that people took with this app, and we manipulated the type of estimate that participants saw. Second, to test whether our results are robust to time durations of different lengths, we additionally manipulated the trip length to be either relatively short (around 45 minutes) or long (around 2 hours and 15 minutes).

Third, in previous studies, we generated the app’s outcomes from a normal distribution where early and late deliveries are equally likely. However, we recognize that in many real-world contexts, such as the domains of driving time estimation and food delivery, right-skewed distributions of outcomes may be more common. There is a limit to how much faster than expected outcomes in these domains can occur (e.g., one can only drive so fast), but bad delays can make outcomes very late. Therefore, in Study 8, we generated

the actual driving durations from log-normal distributions (with a right-skewed shape). Specifically, we generated 50 numbers from a log-normal distribution with shape parameters $\mu = 3.8$ and $\sigma = 0.09$ (short trip conditions; $M = 45$; $SD = 4$) or $\mu = 4.9$ and $\sigma = 0.06$ (long trip conditions; $M = 135$; $SD = 8$). As in previous studies, the actual driving durations of the 20 trips (i.e., how long it took each driver to arrive at their destination) were randomly selected from these 50 numbers.

Manipulation. Participants in this study were randomly assigned to one of six conditions in a 2 (trip duration: short versus long) \times 3 (estimate format: accurate point estimate versus conservative point estimate versus range estimate) between-subjects design. The accurate point estimate corresponded to the mean of the underlying distribution, whereas the range estimate corresponded to the 80% CI and the conservative point estimate corresponded to the upper bound of the range. Hence, in the short trip conditions, the accurate point estimate was “45 minutes,” the conservative point estimate was “50 minutes,” and the range estimate was “40–50 minutes.” And in the long trip conditions, the accurate point estimate was “2 hours and 15 minutes,” the conservative point estimate was “2 hours and 25 minutes,” and the range estimate was “2 hours and 5 minutes to 2 hours and 25 minutes.” Note that because the distribution in the short trip condition had a standard deviation of 4 and the distribution in the long trip condition had a standard deviation of 8, and the range estimate corresponded to the 80% CI in either condition, the range encompassed 10 minutes in the short trip condition and 20 minutes in the long trip condition.

Results and Discussion

Analysis Plan. We preregistered to compare the three estimate format conditions to each other by regressing the dependent measure on condition dummies. Table 5 presents the results. Our discussion below focuses on examining whether the results we previously observed in the food delivery context regarding participants’ preferences for (a) ranges over point estimates and (b) conservative over accurate estimates persist in the driving context.

Point Estimate vs. Range Estimate. Consistent with results in the food delivery context, participants liked the GPS app more when it provided ranges rather than accurate point estimates, $b = 0.39$, $SE = 0.10$, $p < 0.001$. This result held for both the short trip ($b = 0.53$, $SE = 0.14$, $p < 0.001$) and the long trip conditions ($b = 0.25$, $SE = 0.13$, $p = 0.062$). This result suggests that people’s preference for ranges over point estimates is not unique to the food delivery context, to a normal distribution of outcomes, or to a particular time duration.

Table 5. Satisfaction Results of Study 8

	Point estimate		Conservative point estimate		Range	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Short trip	5.13	1.21	4.83	1.33	5.66	0.97
Long trip	5.25	1.10	4.82	1.23	5.50	1.03
All data	5.19	1.15	4.82	1.28	5.58	1.01
	Effect of conservative point (vs. point estimate)		Effect of range (vs. point estimate)		Effect of range (vs. conservative point estimate)	
Short trip	$b = -0.30, SE = 0.14, p = 0.028$		$b = 0.53, SE = 0.14, p < 0.001$		$b = 0.83, SE = 0.14, p < 0.001$	
Long trip	$b = -0.43, SE = 0.13, p = 0.001$		$b = 0.25, SE = 0.13, p = 0.062$		$b = 0.68, SE = 0.13, p < 0.001$	
All data	$b = -0.37, SE = 0.09, p < 0.001$		$b = 0.39, SE = 0.10, p < 0.001$		$b = 0.75, SE = 0.09, p < 0.001$	

Notes. 1. In the Short Trip condition, the Point Estimate was 45 minutes, the Conservative Point Estimate was 50 minutes, and the Range was 40–50 minutes; in the Long Trip condition, the Point Estimate was 2 hours and 15 minutes, the Conservative Point Estimate was 2 hours and 25 minutes, and the Range was 2 hours and 5 minutes to 2 hours and 25 minutes. 2. The regression analyses in the “All Data” row included fixed effects for the trip duration condition.

Accurate Estimate vs. Conservative Estimate. In contrast to the results in the food delivery context, participants did not prefer conservative point estimates to accurate point estimates in the driving context. In fact, participants liked the GPS navigation app *less* when the app provided conservative point estimates that overestimated the trip duration rather than accurate point estimates, $b = -0.37, SE = 0.09, p < 0.001$. These results held in both the short trip ($b = -0.30, SE = 0.14, p = 0.028$) and the long trip ($b = -0.43, SE = 0.13, p = 0.001$) conditions.⁸

Study 8 once more highlights that companies can benefit from providing ranges instead of point estimates, as range estimates seem to be liked more. Importantly, Study 8 also finds that while people may prefer point estimates at different locations in different domains, their preference for ranges over point estimates is robust to different domains (food delivery and GPS), different time durations, and different underlying outcome distributions. Taken together, these results suggest that companies operating in inherently uncertain contexts may be better off by providing customers with ranges rather than point estimates when customers can evaluate how the outcomes align with the time estimates.

General Discussion

We find that when evaluating time estimates, people are more satisfied with apps that provide ranges rather than point estimates. This preference holds across different domains (food delivery and GPS), different lengths of durations, and different underlying outcome distributions. Consistent with this, in an incentive-compatible design where time estimates are consequential, participants were more likely to choose a model providing range estimates than one providing point estimates after seeing both models perform. Moreover, we find that participants liked ranges more than conservative point estimates equivalent to the upper bound of the range, suggesting that greater customer satisfaction for ranges

is not merely driven by a distaste for outcomes falling beyond the upper bound of an estimate.

Importantly, we also find evidence for a relevant boundary condition of the effectiveness of ranges: ranges are no longer favored when they are excessively wide. Widening a range presents a trade-off between accuracy and informativeness (Yaniv and Foster 1995). The results from our studies suggest that people pay attention to this trade-off, and that they seem to only prefer ranges to point estimates as long as the accuracy gained from providing (or widening) a range does not lead to too much information being lost. This becomes particularly evident when we analyze the accuracy measures collected in Studies 1–3 and 6–8 (see Supplement 2). Across all of these studies, participants consistently recognized that ranges are more accurate than point estimates, and this includes the extremely wide ranges that we presented to participants in Study 2. Importantly, however, even though participants correctly recognized that extremely wide ranges are at least equally likely to contain the actual outcomes as more moderate ranges, they did not like them as much as moderate ranges or even point estimates. This suggests that participants correctly recognized that extremely wide ranges are not informative. We present all details of these analyses of the accuracy measures in Supplement 2.

We propose that people prefer range estimates compared with point estimates, because range estimates lead people to expect greater variability, thus rendering different outcomes less likely to violate people’s expectations. In line with this proposal, participants in our studies constructed wider distributions of potential outcomes when given a range estimate compared with when given a point estimate. They also perceived more outcomes as consistent with a range estimate than a point estimate, and less frequently judged outcomes as violating their expectations when provided a range estimate rather than a point estimate.

Theoretical Contributions

Our findings make several contributions to the literature. First, we contribute to prior work on communicating uncertainty to people, which represents a major effort in the recent literature. Past research has examined how people perceive uncertain estimates in product information, medical advice, financial forecasts, and other domains (e.g., Thomas et al. 2010; Du et al. 2011; Xie and Kronrod 2012; Zhang and Schwarz 2012; Gaertig and Simmons 2018, 2023; Howe et al. 2019). We contribute to this body of research and find that in the domain of time, a commonly experienced yet understudied domain, uncertain estimates in the form of ranges are not disliked—they are actually preferred.

Second, we shed light on the process underlying people's preference for uncertain estimates in the domain of time. Our findings largely align with Yaniv and Foster's (1995) proposal that both accuracy and informativeness are important determinants of people's evaluation of estimates under uncertainty. However, beyond that, we also investigate and find that compared with point estimates, range estimates lead people to expect greater variability in future outcomes. In Studies S2 and S3, we provide evidence for this expectation formation. And in Study 5, we show that range estimates make outcomes less likely to violate people's expectations and that this perception relates to people's satisfaction with an app.

Moreover, the current findings also deepen our understanding of how people incorporate judgments of accuracy and informativeness when evaluating time estimates. Most notably, Yaniv and Foster's (1995) model assumes that estimation errors of the same magnitude are equally costly regardless of the direction. In contrast, our work suggests that this is often not the case in the domain of time, in which being five minutes early and being five minutes late can be differentially costly. This is exemplified in the results of Studies 6 and 7, in which people were either indifferent between accurate estimates and conservative estimates (which are strictly less accurate than accurate estimates while being equally informative) or preferred conservative estimates. These results violate the assumption in Yaniv and Foster's (1995) model and underscore the importance of examining people's preferences for uncertain estimates within the specific domain of interest.

Third, our findings also contribute to the broader literature on time perception. Extant literature has shown that people's time perception is malleable and can be affected by many contextual factors. For example, people judge the length of a time interval differently depending on whether the duration or the date is highlighted (LeBoeuf 2006, LeBoeuf and Shafir 2009, Munichor and LeBoeuf 2018). In addition, people tend to mentally represent events in time in a categorical way (e.g., Tu and Soman 2014, Tonietto et al. 2019, Donnelly et al. 2022), leading them, for example, to judge time

intervals that span category boundaries as longer than those that do not (Donnelly et al. 2022). Adding to this work, our findings in Studies S2 and S3 suggest that people form different expectations about the variability of future outcomes given different types of time estimates. We find that compared with point estimates, time estimates in the form of ranges lead people to anticipate greater variability in future outcomes.

Managerial Implications

Our research has important practical implications as well. Companies offering time estimates to customers face significant challenges with respect to how to best communicate their estimates given how unpredictable future outcomes are. Our research offers a simple suggestion: when customers can experience outcomes, it seems to be almost always better to communicate the uncertainty inherent in the outcomes by providing a range rather than a point estimate. In fact, providing a range alone is even more preferred than providing a range and a point estimate. This suggestion, however, comes with a cautionary note: the range cannot be too wide to the point that it is no longer informative.

Recognizing the important role of outcome information has implications for increasing long-term customer satisfaction. Most existing customers engage with the platform over time and learn how well the estimates provided by the platform perform (e.g., when placing multiple food orders or navigating multiple trips with the same app). Thus, although companies could possibly benefit from making estimates more informative by providing a point estimate the first time a customer uses their service, this strategy may prove costly in the likely event that the customer's expectations are violated.

Future Directions

In addition to offering novel insights into how people evaluate uncertain time estimates, the current research also presents exciting opportunities for future research. First, our theorizing alludes to potential moderators that are worth exploring in future research. For example, while the current research focused on time estimate domains that are inherently uncertain, it is possible that in domains with little or no uncertainty—such as public transportation services that operate on well-established routes and fixed schedules—consumers may prefer point estimates to ranges. In this case, point estimates may be able to increase informativeness without meaningfully decreasing accuracy. Additionally, when outcomes are not immediately available—for example, when customers have to decide whether to place an order upon observing the initial time estimate with no past experience—people will likely base their evaluations on factors other than accuracy. Factors such as the estimate's informativeness, their familiarity with the format, or what kind of estimate they expect to receive in

that particular context may play a more prominent role, and ranges may not be strictly favored. Moreover, although we replicate our results across domains with diverse sources of uncertainty—food delivery, driving, and participants' own task completion—it remains possible that the nature of uncertainty (e.g., external or internal to the service provider) may influence customers' preferences in real-world scenarios. We look forward to future research investigating these possibilities.

In addition, our studies specifically focused on the effect of communicating uncertainty by comparing range estimates to point estimates. However, it would be interesting to consider other ways of communicating uncertainty. For example, companies may provide a margin of error around a point estimate (e.g., “40 +/- 5 minutes”), give round numbers instead of precise numbers (e.g., “40 minutes” instead of “41 minutes”), or add verbal qualifiers to the numeric estimates (e.g., “roughly 40 minutes,” “approximately 40 minutes”). Whether consumers prefer these other communications of uncertainty to receiving a point estimate may depend on the extent to which these communications of uncertainty expand people's expected distributions of outcomes.

Future research could also investigate the downstream consequences of providing precise versus uncertain estimates that could have important implications for customer relationship management. For instance, does the format of time estimates affect customers' fairness perceptions, churn, usage rates, and other important outcomes? It could be the case that consistently providing customers with point estimates and then violating their expectations deteriorates their trust in the service platform.

Another future direction pertains to market competition among different apps providing time estimates. Consider the food delivery industry as an example, where customers may check delivery estimates from multiple services before placing an order. If these services compete based on the estimates that they give, it might be tempting to give overly precise estimates to attract customers. However, this could also backfire once outcomes are realized if customers' expectations are routinely violated. Future work could investigate how best to construct time estimates to strike the balance between customer acquisition and customer satisfaction.

Finally, future research could extend the current results beyond static estimates to dynamic estimates. In real life, ride-sharing apps like Uber and Lyft update estimated arrival times based on real-time traffic information, and food delivery apps like Uber Eats and DoorDash update the estimated delivery time of an order based on preparation time, pickup time, and local traffic. We look forward to future research that investigates how these estimates should be dynamically updated to increase customer satisfaction.

Conclusion

Understanding how communicating inherent uncertainty affects customer satisfaction is a pressing question for companies operating in many different domains. The current research elucidates an important aspect of this decision. We find converging evidence that providing a time estimate as a range leads to greater customer satisfaction than providing a point estimate. Our research contributes to a growing literature on communicating uncertainty and offers feasible solutions to practitioners providing time estimates to customers.

Acknowledgments

The authors thank participants in lab meetings at the University of Pennsylvania and the University of California, Berkeley for their helpful suggestions and Jake Flancer, Jordyn Schor, Gregoria Serretta Fiorentino, and Zaiying Yang for valuable research assistance.

Endnotes

¹ In the current research, we are chiefly interested in the comparison between point estimates and range estimates. Of course, there are other ways to express uncertainty in time estimates, for instance, by giving a round number instead of a precise number (e.g., “40 minutes” instead of “41 minutes”) or adding a verbal qualifier (e.g., “roughly 40 minutes”). We discuss other approaches to convey uncertainty and other time estimate formats used by practitioners in the General Discussion.

² It is also worth noting that Yaniv and Foster's (1995) model “assumed equal penalty for over- and underestimation of the true answers” (p. 431). Although this assumption may be reasonable for general knowledge questions, Yaniv and Foster (1995) acknowledged in their General Discussion that this might not hold true in the domain of time, in which being five minutes early and being five minutes late often have different consequences. Indeed, in our Studies 6 and 7, we find that, in violation of Yaniv and Foster's (1995) model, people may sometimes favor receiving a conservative point estimate that overestimates outcomes to an accurate point estimate that is equally informative. We further discuss this when presenting the results for Studies 6 and 7 and in the General Discussion.

³ If outcomes are normally distributed, 90% of the outcomes fall within 1.645 standard deviations from the mean, and 99% of the outcomes fall within 2.576 standard deviations from the mean. Thus, expanding a range that includes 90% of outcomes to a range that includes 99% of outcomes would widen the range by approximately 57%.

⁴ Participants in all studies reported their age in an open-ended text box. We noticed that one participant in Study 1 indicated their age as 5 years, one participant in Study 2 as 0.36 years, and one participant in Study 7 as 229 years. We suspect that these numbers arose due to typos and therefore excluded the ages of these participants for our calculation of participants' average age in the respective study.

⁵ In our theoretical framework, we used the term “informativeness” to refer to the extent to which the estimate communicates a specific outcome (Yaniv and Foster 1995). It is important to note, however, that we don't expect laypeople to share the same interpretation when responding to the informativeness item that we included in our satisfaction scale. Consistent with this proposal, Supplement 8 presents evidence that participants interpreted this question as asking about the lay meaning of informativeness.

⁶ Note that we pre-tested the completion time for pages with 5–30 sliders, but for the main study, we asked participants to complete 8, 12, and 18 sliders in order to keep the survey at a reasonable length and to ensure that the completion times for the three pages are sufficiently different.

⁷ In Study 7, in addition to the above analysis, we also preregistered to regress each measure on (1) the accurate point estimate condition, (2) the conservative point estimate condition, and (3) the accurate range condition. This allowed us to compare these conditions to the conservative range condition. We provide the results for these comparisons in Supplement 7.

⁸ It is worth noting that apart from the context difference (driving instead of food delivery), we also used a different underlying distribution (log-normal instead of normal) to generate the outcomes. However, we do not expect the difference in distribution to account for the different results between the food delivery and driving contexts. If it did, then we would expect to see the opposite result. The log-normal distribution is more right-skewed than the normal distribution and contains more “late” outcomes. Thus, if anything, we would expect a conservative estimate to feel more accurate and be even more appealing given a log-normal distribution, but we observed the opposite result.

References

- André Q (2016) distBuilder. *Zenodo*, <https://doi.org/10.5281/zenodo.166736>.
- Dieckmann NF, Robert M, Slovic P (2010) The effects of presenting imprecise probabilities in intelligence forecasts. *Risk Anal. Internat. J.* 30(6):987–1001.
- Donnelly K, Compiani G, Evers ER (2022) Time periods feel longer when they span more category boundaries: Evidence from the laboratory and the field. *J. Marketing Res.* 59(4):821–839.
- Du N, Budescu DV, Shelly MK, Omer TC (2011) The appeal of vague financial forecasts. *Organ. Behav. Human Decision Processes* 114(2):179–189.
- Gaertig C, Simmons JP (2018) Do people inherently dislike uncertain advice? *Psych. Sci.* 29(4):504–520.
- Gaertig C, Simmons JP (2023) Are people more or less likely to follow advice that is accompanied by a confidence interval? *J. Experiment. Psych. Gen.* 152(7):2008–2025.
- Grice HP (1975) *Logic and conversation*. Speech Acts (Brill, Leiden, Netherlands), 41–58.
- Hayward MLA, Fitzma MA (2017) Pseudo-precision? Precise forecasts and impression management in managerial earnings forecasts. *Acad. Management J.* 60(3):1094–1116.
- Howe LC, MacInnis B, Krosnick JA, Markowitz EM, Socolow R (2019) Acknowledging uncertainty impacts public acceptance of climate scientists’ predictions. *Nature Climate Change* 9(11):863–867.
- Hu B, Simmons JP (2024) Different methods elicit different belief distributions. *J. Experiment. Psych. General*. Advance online publication, <https://doi.org/10.1037/xge0001655>.
- Jerez-Fernandez A, Angulo AN, Oppenheimer DM (2014) Show me the numbers: Precision as a cue to others’ confidence. *Psych. Sci.* 25(2):633–635.
- Joslyn SL, LeClerc JE (2012) Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *J. Experiment. Psych. Appl.* 18(1):126.
- LeBoeuf RA (2006) Discount rates for time vs. dates: The sensitivity of discounting to time-interval description. *J. Marketing Res.* 43(1):59–72.
- LeBoeuf RA, Shafir E (2009) Anchoring on the “here” and “now” in time and distance judgments. *J. Experiment. Psych. Learn. Memory Cognition* 35(1):81.
- Munichor N, LeBoeuf RA (2018) The influence of time-interval descriptions on goal-pursuit decisions. *J. Marketing Res.* 55(2):291–303.
- Pena-Marin J, Wu R (2019) Disconfirming expectations: Incorrect imprecise (vs. precise) estimates increase source trustworthiness and consumer loyalty. *J. Consumer Psych.* 29(4):623–641.
- Price PC, Stone ER (2004) Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *J. Behav. Decision Making* 17(1):39–57.
- Radzevick JR, Moore DA (2011) Competing to be certain (but wrong): Market dynamics and excessive confidence in judgment. *Management Sci.* 57(1):93–106.
- Sah S, Moore DA, MacCoun RJ (2013) Cheap talk and credibility: The consequences of confidence and accuracy on advisor credibility and persuasiveness. *Organ. Behav. Human Decision Processes* 121(2):246–255.
- Teigen KH, Nikolaisen MI (2009) Incorrect estimates and false reports: How framing modifies truth. *Thinking Reasoning* 15(3):268–293.
- Tenney ER, Spellman BA, MacCoun RJ (2008) The benefits of knowing what you know (and what you don’t): How calibration affects credibility. *J. Experiment. Soc. Psych.* 44(5):1368–1375.
- Thomas M, Simon DH, Kadiyali V (2010) The price precision effect: Evidence from laboratory and market data. *Marketing Sci.* 29(1):175–190.
- Tonietto GN, Malkoc SA, Nowlis SM (2019) When an hour feels shorter: Future boundary tasks alter consumption by contracting time. *J. Consumer Res.* 45(5):1085–1102.
- Tu Y, Soman D (2014) The categorization of time and its impact on task initiation. *J. Consumer Res.* 41(3):810–822.
- Van Der Bles AM, van der Linden S, Freeman ALJ, Spiegelhalter DJ (2020) The effects of communicating uncertainty on public trust in facts and numbers. *Proc. Natl. Acad. Sci. USA* 117(14):7672–7683.
- Xie G, Kronrod A (2012) Is the devil in the details? The signaling effect of numerical precision in environmental advertising claims. *J. Advert.* 41(4):103–117.
- Yaniv I, Foster DP (1995) Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *J. Experiment. Psych. General* 124(4):424.
- Zhang YC, Schwarz N (2012) How and why 1 year differs from 365 days: A conversational logic analysis of inferences from the granularity of quantitative expressions. *J. Consumer Res.* 39(2):248–259.