Optimal Prototyping with Noisy Measurements

Sreekumar Bhaskaran[†] Sanjiv Erat[‡]

November 6, 2024

Abstract

Problem Definition: Prototyping and testing are an integral part of almost any new product development process, helping firms navigate the inherent uncertainties of creating new products. Recent developments in rapid prototyping, including technologies that enable cheaper low fidelity tests, have opened up the possibilities for firms in reconfiguring their product development processes. Firms can, by choosing the level of evaluation fidelity, alter the traditional cost-quality tradeoffs inherent in sequential prototyping. **Methodology and Results:** The current article formulates a general model of sequential search where firms can proceed by obtaining noisy low-fidelity evaluations of their prototypes. Our results demonstrate that the imperfect fidelity of evaluations alters the firm's optimal experimentation, with the starkest difference being that it may make it optimal for the firm to select and launch a prototype that did not yield the best evaluation. In addition, our analysis of optimal measurement technology reveals that the focal firm should demand most precise measurements when their ex-ante uncertainty is moderate (not too high or low). We also consider extensions analyzing how the optimal choice of evaluation fidelity is affected by the number of available prototypes, by operational flexibility (to dynamically change measurement technology), and by the ability to outsource evaluations to an experimentation platform. **Managerial Insights:** We develop managerial insights for how the optimal choice of fidelity and the optimal length of the evaluation cycle should be planned depending on the evaluation costs and the firm's ex-ante uncertainty. The resulting framework offers guidance to product and software development firms to successfully leverage imperfect fidelity experiments.

Keywords: testing, fidelity, search, experimentation platforms, outsourcing

[†]Cox School of Business, SMU. email: sbhaskar@mail.cox.smu.edu
[‡]Rady School of Management, UCSD. email: serat@ucsd.edu

1 Introduction

Traditional product development processes employ build-test cycles as a means to validate and refine the product's design and functionality. These cycles consist of iterative stages where prototypes or early versions of a product are built, assessed, and then refined based on feedback and insights gained during the evaluation phase (Thomke, 1998). An important development in this field has been the emergence of rapid prototyping technologies, which have fundamentally transformed product development and testing cycles across various practical applications. These technologies, while often permitting only lower fidelity tests, have altered the affordability and cost-effectiveness of test cycles (Jacobs, 1992; Pham and Gault, 1998). For instance, in UX design, firms are increasingly relying on lower fidelity evaluations to quickly and effectively obtain user feedback and to incorporate these into subsequent iterations (Busche, 2014). Similarly, engineering firms are employing lower-fidelity simulations to assess product performance and user responses under different conditions. This approach allows them to detect design flaws early in the process, make quick adjustments, and ultimately reduce development costs (Joyner et al., 2022).

Recent changes in how product development teams are structured within organizations have also influenced prototype evaluations and the manner in which they are used in the development process. Traditionally, product development teams, comprised of both design and test engineers, co-existed within a single unified organization and collaborated to move the product from concept to market. However, firms have recently adopted a modified approach to managing the design, prototyping, and testing stages. In this new approach, firms continue to handle the decisions related to design and prototyping, while outsourcing only the evaluation of these prototypes to external platforms. This shift in the workflow, facilitated by the rise of specialized experimentation platforms like Optimizely, Statsig, Applitools, and Appen, also increases the importance of being able to manage lower fidelity evaluations in the firm's test cycle. Specifically, such decoupling creates both a lack of visibility into the actual evaluation process and ambiguity in the product development firm's requirements, and leads to potentially noisy evaluations, raising concerns about the fidelity and accuracy of results (see Jacobs, 1992; Pham and Gault, 1998).

The use of such low fidelity evaluations comes with its own unique set of challenges. Even beyond the question of how to organize test cycles with low fidelity evaluations, the firm often has the ability to influence the "level" of evaluation fidelity. As an example, in UX design, by employing more testers, the firm can obtain more accurate assessments, but at an increased costs. Indeed, this additional decision - how accurately to evaluate the design - is even more significant in settings where the evaluations are outsourced to experimentation platforms. Specifically, in these settings, the firms can request, and the platform can provide higher fidelity evaluations. However, this choice comes at an increased

cost and could even be dependent on the type of prototype being evaluated. As we present in the following example, balancing evaluation quality against the prototyping costs is essential to successfully exploiting the value of low fidelity evaluations.

Low Fidelity Experimentation: An Illustrative Example

To illustrate the unique challenges involved in low fidelity evaluation, consider a specific instance of a product development firm who, looking to build a solution that enables effective interactions with their online customers, engages an experimentation platform that we refer to as Alpha.[1] When considering the design of a solution, the product development firm has the flexibility to explore various approaches. One potential avenue involves crafting a straightforward interface capable of accommodating diverse UX designs, while an alternative path entails the development of intelligent chatbots. However, it may be observed that these two approaches, while geared toward accomplishing the same business goal, rely on distinct methodologies. In the first approach, the firm, using traditional UX design principles, creates interfaces where the customers select from a list of predetermined options (dropdown menus, radio buttons, etc.). In the second, the firm, perhaps by leveraging generative AI techniques, creates chatbots that allow richer (and unstructured) conversations with customers.

The two approaches outlined above, in addition to (any) differences in value and risk, are also distinct in terms of the cost and the fidelity of evaluations that may be obtained. Specifically, as traditional UX is a (relatively) well established area and the focal firm's requirements are unambiguous, evaluating the quality of a UX design is easier, and might even be feasible with inexpensive measurement technologies (such as automated test suites). In contrast, testing a chatbot is prone to (subjective) errors as a result of possible ambiguity in the product development firm's requirements/use-cases, and because of the particulars of the measurement technology (i.e., human testers). These challenges make the evaluation of the chatbot quality (compared to traditional UX quality) likely to be of lower *fidelity*. Thus, the focal firm, when pursuing the traditional UX approach, faces a different set of cost-fidelity tradeoffs compared to the case where they pursue AI chatbots as their preferred approach.

The current study, in a bid to understand these trade-offs, examines the following questions:

1. How do the prototype-specific and measurement technology characteristics, including the inherent ex-ante uncertainty of the prototype and the fidelity/costs of evaluations, affect the firm's prototyping choices?

2. Which measurement technology should the firm employ, and how does the answer depend on the prototype-specific characteristics?

[1] Alpha is a pseudonym for one of the widely used experimentation platforms that motivated the current study. The name, at the request of the platform's management team, has been anonymized to retain confidentiality.

To answer these questions, we begin by formulating a model of "noisy sequential search" where the firm selects the prototype (to be tested), and obtains an evaluation of the prototype's value. These evaluations are assumed to be a noisy signal of the "true value" of the prototype. Moreover, the "size" of this noise depends on the measurement technology that is used (with lower/higher fidelity measurement technologies costing less/more). Subsequently, we solve the model to obtain the firm's optimal actions - which prototypes to evaluate and at what fidelity, and when to stop the evaluations. Next, we also extend our analysis by considering the experimentation platform's problem, and we characterize the test fidelities that they would offer, and the corresponding fees they charge.

The current study makes two key contributions: First, from the methodological perspective we formulate and solve the first, to the best of our knowledge, fully general model of sequential search with noisy measurements. Second, from the substantive perspective, we characterize the optimal policy when the measurement fidelity can be chosen, and we identify the contingencies that would call for choosing higher or lower measurement fidelity in the test cycle.

As the firm's problem, when the measurement fidelity is perfect, is identical to the influential Pandora's box model of sequential search (Weitzman, 1979), our solution of the problem when evaluation fidelity is imperfect (and exogeneous) is a full generalization of Weitzman's model of sequential search. However, despite the additional complexity introduced by imperfect measurements, our analysis yields a surprising conclusion - the optimal policy in this setting is a *fidelity-adjusted reservation price policy*, where the reservation price for each alternative is an index that depends only on the evaluation- and alternative-specific characteristics (à la Weitzman, 1979).

Our analysis of the optimal policy also shows that while the firm, when provided with perfect fidelity measurements, should always select the prototype with the highest evaluation, the same is not true when the evaluations are imperfect. Intuitively, when evaluations are imperfect, the measurements should always be "discounted" (to reflect the possibility of measurement errors). Consequently, the alternative with the highest measurement might not be the same as the alternative with the highest "discounted" measurement. Indeed, our results demonstrate that, when the fidelity is poorer, it is more likely that the firm optimally picks a prototype that did *not* yield the highest measurement.

When the firm can choose specific evaluation fidelities, our results demonstrate that the optimal choice of evaluation fidelity is *non-monotone* in the ex-ante uncertainty of the prototypes. Specifically, for very low uncertainty, the firm finds limited value in high fidelity evaluations (since the true value is most likely close to the expected value), and hence they can save on evaluation costs by employing a lower fidelity evaluation. In contrast, when uncertainty is moderate, the firm requires higher fidelity evaluations so as to decide whether to accept or continue evaluations. Still, for very high values of intrinsic uncertainty, the true value is most likely to be very far away from expected values, and for

these settings, the optimal decision, irrespective of the measurement fidelity, is fairly easy (with very poor/good values implying rejection/acceptance). Thus, when the intrinsic uncertainty is very high, perfect fidelity measurements are once again of limited utility, and the firm saves on their costs by employing low fidelity measurement.

We extend our analysis to situations where the number of available design alternatives is more limited. As very large number of available alternatives guarantees that at least one is exceptional and will stand out (irrespective of the evaluation fidelity), we find that it is with moderate number of alternatives that the firm employs the highest fidelity evaluation. In the case where the firm outsources the evaluation to an experimentation platform, we also examine the platform's choice of offered fidelity and the associated fees. Results mimic our findings in the inhouse evaluation case, and we show that the platform offered evaluation fidelity is also *non-monotone* in the uncertainty of the prototypes. Moreover, since the focal firm has greater upside potential with increasing ex-ante uncertainty, the platform charges a higher fee for their evaluations when the prototypes have greater uncertainty. Taken together, these provide an interesting result: when the ex-ante uncertainty of the design alternatives increases from moderate to high values, the platform should offer lower fidelity evaluations, but for a higher fee.

The rest of the article is organized as follows: We provide a selective review of related product development and innovation literature in §2. We begin by offering a general dynamic model of decentralized low-fidelity evaluations of prototypes in §3, and solve the model to fully characterize the firm's optimal dynamic policy. Subsequently, a more concrete operationalization of evaluation noise is offered, and the determinants of firm's optimal decisions is examined in §4. Finally, §5 discusses various extensions including the case of dynamic choice of evaluation fidelity (§5.1), the effect of size of design pool (§5.2), and experimentation platform's optimal choices (§5.3). §6 concludes.

2 Review of Related Literature

There are two streams of research that relate to the current study: (i) the product development literature that examines experimentation, and (ii) the methodological search theory literature that provides the formal machinery for testing and prototyping models. We briefly discuss each of these below.

2.1 Testing & prototyping in NPD

The recognition that experimentation is a critical part of the product development process has motivated many of the early studies of prototyping and testing processes (see for instance, Thomke, 2003).

A dominant theme within this research stream centers around the operationalization of a firm's test schedule, and specifically the question of how many prototypes should be tested and how they should be selected or discarded (Thomke, 1998; Dahan and Mendelson, 2001). When it comes to sequential test schedules, beyond the cost savings that can come from earlier termination, the iterative process may also exploit "relatedness" between prototypes and enable the firm to learn from one prototype to the next, which can then enhance value (see, Erat and Kavadias, 2008).

While some of the conceptual literature from this stream have argued for the value of lower fidelity tests, and more broadly the accuracy vs efficiency tradeoff that firms should manage when choosing the mode of evaluations (see for instance, Kornish and Hutchison-Krupat, 2017), the theoretical literature and formal models have primarily focused on optimizing a firm's experimentation cycles, often assuming that evaluations will always accurately reveal the true values of the final product. An early study of the role of fidelity in NPD is offered by Loch et al. (2001), who while restricting their attention to the relative value of prototypes, did consider the possibility that a test might not always accurately signal the true value of a product. In specific, they assume that measurement reveals, possibly imperfectly, whether a given prototype is the "best" among all the available (tested and untested) alternatives. Still, from a conceptual standpoint, Loch et al. (2001) view the measurement fidelity as less of a choice, and more as an intrinsic property determined by nature. Our study, motivated by many practical experimental methodologies, including, more recently, A/B testing of prototypes (see for instance, Berman and Van den Bulte, 2022; Koning et al., 2022; Bhat et al., 2020), work under the premise that testing reveals a more or less precise signal of a prototype's true value depending on whether we devote greater or fewer resources toward the test. Thus, the current study expands the scope of noisy experimentation by viewing the measurement fidelity itself as a choice.

A related stream of delegated experimentation literature that examines settings where the evaluation is outsourced has considered models which view measurement fidelity as an exogeneous parameter (see for instance, Gerardi and Maestri, 2012; Schlapp and Schumacher, 2022). While details differ, one common finding relates to the importance of incentive contracts in managing agency issues (including issues connected to truthful reporting of evaluations and effort provisioning). Our setting, by contrast, is meant to emphasize measurement fidelity as a choice (and not just as a parameter), and to understand its role (irrespective of whether the fidelity issues emerge because of outsourced evaluations or from intrinsic characteristics of the measurement technology).[2] This focus allows a straightforward and complete generalization of Weitzman (1979), and we show that a "fidelity-adjusted reservation price policy" still proves optimal.

[2]Classic agency problems, such as deceptive reporting or effort shirking, especially for an experimentation platform attempting to obtain repeat business from several clients, is less likely as it can very easily creates long-term reputational risks.

The theme of "agency issues" and associated distortions have found increasing prominence in product development, particularly as the complexity of projects and the escalating costs forced firms to engage in collaborative ventures with external entities (Bhaskaran and Krishnan, 2009). Studies within this domain examined strategies that firms should adopt in such collaborations, encompassing aspects such as design, development, and, in certain instances, post-development commercialization (Roels et al., 2010; Crama et al., 2017; Rahmani et al., 2017; Tian et al., 2021). Recent research has expanded the analysis to consider the context of testing and prototyping, and the agency-related challenges that arise when these functions are outsourced. Terwiesch and Xu (2008) consider the case of crowdsourcing innovations from outside parties. Rahmani and Ramachandran (2021) considers the benefit of ex-ante commitment to unambiguous rules in aligning incentives in a delegated search process. In a model of search without recall, Zorc et al. (2023) examine the optimal contracts under delegation and the conditions that determine the choice between in-house vs delegated search. In contrast to much of this literature, the current article has as its primary goal the consideration of measurement fidelity, rather than the agency issues that arise when the entirety of the prototyping process is delegated. Specifically, we examine a scenario where *only* the assessment of the prototype performance is delegated to an outside experimentation platform, while the firm retains control over key prototyping decisions. Thus, after receiving the platform's evaluation, the firm independently decides how to proceed, including which prototype to test next and when to stop testing.

2.2 Methodology/Search theory literature

In a seminal contribution to search theory, Weitzman (1979) formulated a model where he considered the case of sequentially searching through alternatives with the goal of choosing the best one. Under the assumption that evaluations are perfect, he demonstrated that the optimal dynamic rule turns out to have a surprisingly simple structure. Specifically, his results show that we can associate an index (a "reservation price") to each alternative (where the index depends only on the characteristics of the alternative), and that the optimal policy at each stage is to look at the alternative with the highest reservation price. Moreover, the agent optimally stops once the best value they have found so far exceeds all remaining reservation prices.

Building on this seminal work, numerous scholars have explored how the optimal policy may be implemented in practice and how it may be impacted when considering various practical aspects of the testing process. Through lab experiments, several researchers have examined the question of how searchers make their choice when sequentially evaluating between multiple alternatives (see for instance, Sommer et al., 2020; Özkan-Seely et al., 2023). In a theory paper, Adam (2001) considers the possibility that firms can learn from one alternative to another, and he demonstrates that characterizing

the optimal reservation price policies is possible only for a limited set of learning models. This research was extended by Erat and Kavadias (2008) who showed that with unrestricted learning (but with restrictions placed on the alternatives themselves), the optimal policy is no longer the single threshold reservation price policy (of Weitzman, 1979). Doval (2018) considered the case where the agent is allowed to select any alternative (whether or not she has looked at it previously), and showed that the reservation price policy is no longer optimal. Thus, extensions to the basic Weitzman model suggest that the model may be "balanced on the edge," where even small changes render reservation-price type policies suboptimal.

Similar to much of the product development literature on experimentation and testing, the above cited studies also make the assumption that fidelity of the evaluations are perfect. In a different context, McCardle (1985) focuses on the choice between a status-quo and a single new uncertain alternative, and examines measurement fidelity and how it affects the trade-off between collecting more information about a new technology versus immediately adopting it. Still, studies in this stream treat fidelity as a parameter, and only view the extent of information aquisition as the experimenter's choice. Our focus, by contrast, treats the evaluation fidelity as a choice and we characterize the experimenter's optimal decisions.

In the related context of decision theory literature, there is a long history of research examining the problem posed by noisy measurements/assessments and possible solutions to mitigate its effects. For instance, in a study examining the so-called "optimizer's curse," Smith and Winkler (2006) argue that positive measurement errors, by making a decision more likely, make post-decision regret inevitable. Kettunen and Salo (2017) extend this insight by showing that random estimation errors, by affecting the project selection, has significant impact on the (downside) risks in project portfolios. Similarly, Bansal et al. (2017) examine approaches to combining noisy expert judgments and demonstrate the practical utility of their approach in an economically significant setting. The current article also considers evaluation fidelity, but in the context of product development processes, and we examine its implications to the operational setting of prototyping.

3 General Model of Outsourced Experimentation

Consider a firm who is developing a new product for which it is currently evaluating different alternative designs. The firm has N distinct prototype designs that it may choose from, and has to evaluate these alternatives before deciding which among those should be chosen for launch. Consistent with motivating examples presented previously, we assume that the measurement technology does not yield perfect evaluations, and that the firm incurs a cost w_i for evaluating prototype i. Moreover, we also

assume that the focal firm may choose "better" measurement technology to improve the precision of these evaluations, albeit at a higher cost.

Given the set of measurement technologies and the corresponding costs they entail, the firm conducts evaluations sequentially, with the ultimate goal of selecting exactly one of them to launch. At every iteration, the firm evaluates an alternative, and then, contingent on the results of the evaluation, chooses whether to continue. When the firm is satisfied (with the results from the evaluations so far), it stops its test cycle, and picks one among the evaluated alternatives to launch. In other words, the firm chooses the fidelity of the measurement technology, may evaluate the prototypes in any order, and choose to stop at any time, and select any one of the already evaluated alternatives (or select none and thus obtain the normalized outside option of 0).

As described above, a key factor that influences a firm's decision with respect to selection of alternatives is the fidelity of the measurement technology that is available for evaluations. Before formalizing the firm's choice of measurement technologies, we first offer below a general framework of sequential search when the evaluations are noisy. Subsequently, by linking the firm's optimal test schedule with the offered measurement technologies, we characterize the optimal choice of evaluation fidelity.

Formulation & Analysis of Sequential Search with imperfect measurements

We build on the prior literature and start with the assumption that prototype i ($i = 1, \cdots, N$) has true value $\mathbf{P_i}$, where $\mathbf{P_i}$ is a random variable with pdf $f_i(P_i)$ (see for instance, Erat and Kavadias, 2008). Because of the specifics of the measurement technology, the evaluations of any prototype i does not fully eliminate the ex-ante uncertainty about its true value. In specific, we assume that the evaluation of a prototype i can only yield a noisy imperfect signal (measurement) $\mathbf{m_i}$.

To understand the effect of noisy evaluations on the firm's optimal testing policy, we start with a fully general model of noisy measurement, where we assume that the true value $\mathbf{P_i}$ and it's corresponding measurement $\mathbf{m_i}$ is given by joint probability density function $f_i(P, m)$. In this conceptualization, the "noisiness" of the measurement, i.e., the extent of test/measurement fidelity, is implicitly incorporated through the joint probability density function $f_i(P, m)$. For instance, if the measurement was noiseless (high fidelity), the true value should be exactly equal to the measurement (i.e., $\Pr(P = x | m = x) = 1$); and with noisy measurement (low fidelity), the true value will (probabilistically) be different from the measurement. More generally, the "spread" of the conditional distribution $f_i(P_i|m_i) = \frac{f_i(P,m)}{\int_{-\infty}^{\infty} f_i(P,m)dm}$ captures the extent to which the signal reduces the uncertainty, and thus this spread may be used to parameterize the fidelity of evaluations. But we shall defer offering an explicit parameterization of the test fidelity till Section 4.2, and for now focus on characterizing how

the firm's optimal experimentation is affected by the mere presence of measurement noise (as opposed to its "size").[3]

The sequence of events in our noisy search model are as follows: Given the set of N alternatives available and the associated costs w_j ($j \in \{1, \cdots, N\}$) for evaluating them, the firm chooses to evaluate one of the alternatives i. The firm (i) conducts the evaluation and obtains the measurement $\mathbf{m_i}$ and (ii) incurs a cost w_i. Subsequently, the firm, factoring in all of the previous evaluations, decides whether to continue evaluation (if any). This continues till the firm stops further evaluations and accepts one of the already tested alternatives.

Next, we offer a full characterization of the optimal dynamic test schedule chosen by the firm. For ease of exposition, define a fidelity-adjusted expected value $Q_i(m)$ and fidelity-adjusted probability density function $g_i(x)$ as follows:

$$Q_i(m) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \frac{f_i(P,m)}{\int_{-\infty}^{\infty} f_i(P,m)\,dm} P dP \tag{1}$$

$$g_i(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_i\left(P, Q^{-1}(x)\right) dP \tag{2}$$

Note that $Q_i(m)$ is the fidelity-adjusted expected value of the i^{th} alternative when we observe an evaluation m; and $g_i(x)$ is the pdf of the post measurement (expected) value from alternative i given the noisy evaluations.

Finally, let R_i be the *fidelity-adjusted reservation price* of i^{th} alternative, defined as the (smallest) solution of the equation $E[\max\{R, \mathbf{X}_i\}] - R = w_i$ where $\mathbf{X}_i \sim g_i(x)$.[4] Reorder the prototypes such that $R_i \geq R_j$ for $i < j$. Let prototype 0 represent the option of not selecting any of the prototypes. Finally, let V_i and B_i be defined as follows for $i = 0, \cdots, N$.

$$V_0 \stackrel{\text{def}}{=} 0 \qquad\qquad B_0 \stackrel{\text{def}}{=} 0$$

$$V_i \stackrel{\text{def}}{=} \max\{V_{i-1}, Q_i(m_i)\} \text{ for } i > 0 \qquad B_i \stackrel{\text{def}}{=} \begin{cases} B_{i-1} & \text{if } V_{i-1} > Q_i(m_i) \\ i & \text{o/w} \end{cases} \text{ for } i > 1$$

Note that V_i tracks the best fidelity-adjusted expected value seen so far; and B_i the corresponding alternative. With the above definitions, Theorem 1 gives the complete characterization of the firm's optimal dynamic policy.

[3] We borrow the terminology from the product development literature to build our model. Still, it is to be noted that the model is a generalization of Weitzman (1979) model, and thus applicable in any situation where the decision-maker employs a noisy measurement technology and sequentially evaluates alternatives with the goal of selecting the best one.

[4] Such a solution exists since $E[\max\{R_i, Y_i\}] - R_i = E[\max\{0, Y_i - R_i\}]$ is non-increasing in R_i, and has the range $(0, \infty)$.

Theorem 1. *The optimal dynamic policy is a reservation price policy where it is optimal to stop and accept prototype B_n after the n^{th} evaluation ($n \in \{0, 1, \cdots, N\}$) if and only if $V_n \geq R_{n+1}$. Otherwise, she should ask for prototype $n+1$ to be evaluated.*

All technical proofs are given in the Appendix. The theorem demonstrates that

1. The order in which the firm undertakes evaluations is predetermined (i.e., is independent of the actual realizations), and is determined by indices (reservation prices) that depend only on the characteristics of a given prototype and the measurement technology (i.e., R_i depends only on the cost incurred w_i, and the joint distributions of $\mathbf{P_i}$ and $\mathbf{m_i}$)

2. The stopping rule is a threshold policy, where the thresholds are the pre-computed reservation prices, and where stopping occurs when (a monotone function $Q(\cdot)$ of) reported measurements exceeds the threshold.

It is useful to note that, despite the extra complications introduced by "noisy" measurements, the optimal policy still proves to be a reservation price policy, although adjusted to account for the fidelity of the evaluations. In other words, the optimal policy is a *fidelity-adjusted reservation price policy.* Moreover, much of the interesting structural properties of the optimal policy are similar to those found in the classic Weitzman (1979) Pandora's box problem. Still, the actual choices that the firm makes (including when to stop and which alternative to evaluate next) depends on the functions $Q_i(\cdot)$ and $g(\cdot)$, both of which implicitly depend on the fidelity of the evaluations (i.e., $f_i(P, m)$, the joint distribution of $\mathbf{P_i}$ and $\mathbf{m_i}$).

Having characterized how the presence of noisy measurements alters the firm's optimal evaluation policy, we now seek to provide a deeper quantitative analysis of how the outcomes, test schedules, and expected costs are affected by the "size" of the noise. To delve deeper into these questions, we need to make specific assumptions about joint distribution of $\mathbf{P_i}$ and $\mathbf{m_i}$ so as to formalize both the degree of precision of the measurement technology, and how the evaluation cost relates to it. In the next section, we do this by considering an important special case of the general model presented above. Further analysis of this special case also allow us to consider settings where the firm determines the optimal evaluation technology for testing their prototypes.

4 Structural Properties of the Optimal Experimentation Policy

To provide more structure to the firm's design space and the measurement technology, assume the following: the "true" value of $\mathbf{P}_i$ of prototype i is distributed normally with mean μ_i and variance σ_i^2. And, the associated measurement is $\mathbf{m}_i = \mathbf{P}_i + \epsilon_i$ where ϵ_i is normally distributed with mean 0

and precision τ_i^2 (i.e., the variance of ϵ_i is $1/\tau_i^2$). We refer to the τ_i parameter as the fidelity of the measurement since a higher (lower) value of τ_i represents less (more) noisy measurements. Note that in this parameterization, the case of perfect measurement occurs when fidelity/precision $\tau = \infty$, and for all finite values of τ, we have imperfect measurements. Furthermore, the above assumptions imply that $[\mathbf{P}_i, \mathbf{m_i}]$ is jointly normal with mean $[\mu_i, \mu_i]$ and covariance matrix $\begin{bmatrix} \sigma_i^2 & \sigma_i^2 \\ \sigma_i^2 & \sigma_i^2 + 1/\tau_i^2 \end{bmatrix}$.

4.1 Optimal Test Cycle and Dependence on Test Fidelity

Given the above specification of the joint pdf $f_i(P, m)$, we may apply Theorem 1 and obtain the optimal dynamic policy. For completeness, we state this below in the next theorem.

Theorem 2. *Let $R(w)$ be the solution of $x = E[\max\{x, \mathbf{Z}\}] - w$ where $\mathbf{Z}$ is the standard normal. Define an index ("fidelity adjusted reservation price") for each alternative i as $R_i \stackrel{def}{=} \mu_i + \frac{\sigma_i}{\sqrt{1+1/(\tau_i\sigma_i)^2}} R\left(w_i \frac{\sqrt{1+1/(\tau_i\sigma_i)^2}}{\sigma_i}\right)$ $(i = 1, \cdots, N)$, and reorder the prototypes in the decreasing order of R_i.*

Then, the optimal dynamic policy after n evaluations is to

- *(i) continue to conduct more evaluations iff $\max_{i \le n} p_i < R_{n+1}$. Upon continuation, the firm should evaluate prototype $n+1$.*
- *(ii) stop evaluations iff $\max_{i \le n} p_i \ge R_{n+1}$. Upon stopping, the firm should choose alternative $k = \arg\max_{i \le n} p_i$*

where $p_0 \stackrel{def}{=} 0$, and for each prototype i that yielded measurement m_i, the fidelity adjusted value $p_i \stackrel{def}{=} \mu_i + (m_i - \mu_i) \frac{1}{1+1/(\tau_i\sigma_i)^2}$.

Theorem 2 illustrates that the optimal dynamic policy is a threshold policy that depends both on fidelity of the evaluations (i.e., τ), and on the alternative-specific characteristics (i.e., its intrinsic uncertainty σ and expected value μ). To further understand the nuanced role of evaluation fidelity on the optimal policy, we now mute other sources of heterogeneity (by setting $\sigma_i = \sigma$ and $\tau_i = \tau$) and then examine how the fidelity of the measurement technology (τ) influences the choice of which prototype will be chosen for launch at the end of the test cycle.

Proposition 1. *For the case where $\sigma_i = \sigma$ and $\tau_i = \tau$, the probability that the firm optimally picks a prototype r over a prototype s when the measurement of prototype s is superior to prototype r is given by $\Psi(\tau) \stackrel{def}{=} \left(\Phi\left((\mu_r - \mu_s)\frac{\sqrt{\sigma^2+1/\tau^2}}{\sqrt{2}\sigma^2}\right) - \Phi\left((\mu_r - \mu_s)\frac{1}{\sqrt{2(\sigma^2+1/\tau^2)}}\right)\right)^+$ where $\Phi(\cdot)$ is the cdf of the standard normal.*

Furthermore, $\Psi(\tau)$ is non-increasing in τ.

Proposition 1, by isolating the effect of imprecision in the measurement technology, illustrates an interesting consequence of noisy measurements on the firm's testing policy. Specifically, it may be seen that upon terminating the evaluation cycle, the firm may find it optimal to pick the prototype that did *not* yield the highest measurement. Indeed, when the measurement technology offers perfect fidelity, a higher measurement always implies higher true value, and consequently, the firm should always pick the prototype that yielded the highest measurement. However, when evaluations are imperfect ($\tau < \infty$), the measurements are always "discounted" (to reflect the possibility of measurement errors). Consequently, the alternative with the highest measurement might not be the same as the alternative with the highest "discounted" measurement.

Stated differently, as long as there is measurement error ($\tau < \infty$), there is always a positive probability that the firm will optimally pick a prototype that did *not* yield the highest measurement. It is worth pointing out this result does not require that a given alternative has higher variance in comparison to others or that its measurement was obtained using comparatively poorer fidelity measurement technology (i.e., this result holds even when $\sigma_i = \sigma$ and $\tau_i = \tau$). Thus, the possibility of noise in evaluations is what requires the firms to alter their testing policy. In addition, the proposition also demonstrates how the likelihood of not picking the alternative yielding the highest measurement varies with the measurement fidelity. Specifically, the mismatch likelihood is larger (smaller) when the measurement technology has lesser (greater) fidelity. Thus, the use of cheaper but lower fidelity tests (see for instance Jacobs, 1992; Pham and Gault, 1998) makes it less likely that the firm stops and selects the prototype that yielded the best measurement.

Theorem 2 also permit us to quantify the effect of other model parameters on both the dynamics of evaluation cycle and its expected duration. These are stated and explained in the propositions 2 and 3 below.

Proposition 2. *R_i, the fidelity adjusted reservation price of alternative i, is*

(i) increasing in its mean μ_i, increasing in its standard deviation σ_i, decreasing in the cost w_i, and
(ii) increasing in the measurement fidelity τ_i.

Consistent with past literature (see for instance, Erat and Kavadias, 2008), fidelity-adjusted reservation price (i) increases with mean (because it offers more value upon stopping), (ii) increases with variance (because it offers a larger potential upside option), and (iii) decreases with evaluation cost (because it makes evaluations less attractive). In addition to confirming that these insights continue to hold, when the evaluation is of higher fidelity, the fidelity-adjusted reservation price increases. Intuitively, with greater measurement fidelity, the value of obtaining the evaluation is higher, and thus increases the corresponding reservation price. A straightforward consequence of this is that the fidelity

of evaluations may alter the optimal sequence of evaluations and stopping decisions. In specific, lower fidelity leads to (lower reservation prices, and thus to) the alternative being deferred to later in the evaluation sequence, or even to be forgone more readily.

Still, does a firm, given the above logic, always ask for fewer evaluations when the measurement technology is of lower fidelity? The answer, it turns out, is more nuanced than suggested by our reasoning above. On the one hand, these evaluations (using lower fidelity measurement technologies) are less informative, and the firm might consequently request fewer of them. On the other hand, the lower fidelity evaluations are not reliable and stopping prematurely might also increase the opportunity cost and thus prompt more evaluations. Which of these effects prove stronger depends on other model parameters. The proposition below identifies the contingencies where we can expect one or the other to dominate.

Proposition 3. *Probability that the firm will request evaluation of the k^{th} alternative is*

$$P_k \overset{def}{=} 1_{\{p_0<R_k\}}\Pi_{j=1}^{k-1}\Phi\left(\frac{\sqrt{\sigma_j^2+1/\tau_j^2}}{\sigma_j^2}\left(R_k-\mu_j\right)\right)$$

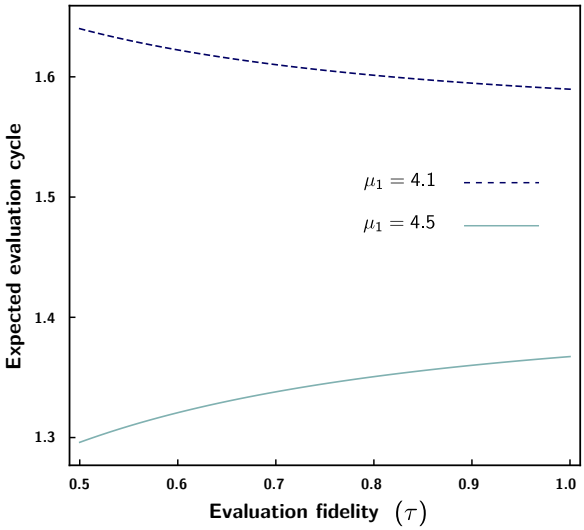
where $\Phi(\cdot)$ is cdf of the standard normal. Furthermore,

(i) P_k (for $k>i$) increases/decreases in measurement fidelity τ_i if μ_i is sufficiently high/low.

(ii) The optimal expected search duration is increasing/decreasing in the measurement fidelity τ_i when μ_i is sufficiently high/low.

Proposition 3 characterizes the likelihood that a particular prototype would be evaluated. Not surprisingly, this probability depends on both the ex-ante expected value of the prototype as well as the fidelity of the measurement technology. What is interesting, however, is that an improvement in measurement fidelity does not impact all the alternatives in an identical manner. Specifically, for an alternative whose intrinsic expected value is lower (as captured by a lower μ_i), an increase in the fidelity of the measurement technology decreases the likelihood of that alternative being evaluated. However, for an alternative with high μ_i, a similar increase would optimally result in a higher likelihood of the alternative being evaluated.

To understand the intuition behind this result, it is useful to recall that when the fidelity is poor (i.e., τ_i is low), the evaluation is less informative and hence it should be discounted more heavily. The rationale in continuing the evaluation process is the hope that an untested prototype could result in a value that is greater than those that have already been tested. When the test fidelity is poor, our ex-ante beliefs play a larger role. Specifically, if the prior μ_i is high, we rely more on our (high) prior rather than the evaluation which only serves to lower the continuation probability (after evaluating



Note: $\mu_2 = 3, \sigma_1 = \sigma_2 = 1, \tau_2 = 1, w_1 = 0.05, w_2 = 0.01$ *in the plotted example with two alternatives.*

Figure 1: Impact of fidelity τ_1 on the duration of the evaluation cycle

alternative i). In contrast, if the prior μ_i is low, then relying on our (low) prior prevents immediate termination of test cycle, and instead increases the likelihood of continuation.

Figure 1 illustrates the contingent relationship between the fidelity of the measurement technology and the expected number of tests that we should plan for. Note that the expected number of alternatives that are tested is increasing in the continuation probability of each alternative.[5] Hence, the expected length of the test cycle will demonstrate identical qualitative behavior to the continuation probability explained earlier. Specifically, when the mean value of the alternative is higher, the expected duration of the evaluation cycle increases with fidelity. In contrast, when the mean value of the prototype is lower, then the reverse holds true, and so, the evaluation cycle might be shorter on expectation when using higher fidelity measurement technology. We next examine how the firm should consider these factors while choosing the level of fidelity that should be employed for the evaluations.

4.2 Choice of Optimal Measurement Technology

So far we have determined the optimal dynamic decisions when the measurement technology available for an alternative i is assumed to have an arbitrary fidelity τ_i, and the corresponding cost of evaluating that alternative is w_i. However, in practice, the focal firm may have the flexibility to choose the specific measurement technology to employ based on their requirements for more or less precise evaluations. As an example, the firm might be willing to employ more testers for their UX design, at higher costs, so as to generate more accurate and higher fidelity assessments. In other instances, a firm might be cost conscious and be willing to work with lower fidelity evaluations. In this subsection, we extend our

[5] Since the expected duration is given by $P_1 + P_2 + \cdots$

model to identify the optimal fidelity in the measurement technology that a firm should choose.

To focus on the optimal choice of the measurement technology, and to eliminate the possibility that heterogeneity of the alternatives (such as different means and variance) is driving our insights, we shall make the additional simplifying assumption that the means and variances are identical across all N prototypes; i.e., $\mu_i = \mu$, $\sigma_i = \sigma$. Given the N alternatives available to evaluate, we begin by analyzing the scenario in which the firm is constrained to choose the corresponding measurement technologies τ_i $(i = 1, \cdots, N)$ *before* starting the test cycle. This "static" choice of test fidelity corresponds to settings where the resource allocation and budgeting decisions necessitate long term planning and hence have to occur prior to any evaluation being conducted. Still, in some situations, it is feasible that the firm might have the capability to dynamically modify their requested fidelities, depending on the observed evaluations. We shall defer the analysis of such contexts till §5.1.

As mentioned previously, the cost of evaluating an alternative w depends on the underlying measurement technology. In specific, we shall assume, without loss of generality, that this cost $w = w(\tau)$ is a non-decreasing function of measurement fidelity.[6] Thus, the firm, prior to doing any evaluation, has the option to choose from different measurement technologies for each alternative, with a choice of higher (lower) fidelity τ corresponding to higher (lower) cost.

The next proposition characterizes, for the full evaluation cycle, the sequence of optimal fidelities.

Proposition 4. *Let* $\theta \stackrel{def}{=} \sqrt{1 + 1/(\tau\sigma)^2}$. *Conditional on asking for evaluation of alternative* i $(i = 1, \cdots, N)$, *let* τ_i^* *be the firm's profit maximizing measurement fidelity. Then,*

(Case A) $\tau_1^* < \tau_2^* < \cdots < \tau_N^*$ *if* $\frac{R(\theta w)}{\theta}$ *is an decreasing function of* τ, *and*
(Case B) $\tau_1^* > \tau_2^* > \cdots > \tau_N^*$ *if* $\frac{R(\theta w)}{\theta}$ *is a increasing function of* τ.

Proposition 4 above is interesting for two reasons. First, it demonstrates that the firm should not, in general, choose the same measurement technology for every alternative, and that they would find it valuable to change their evaluation fidelity requests based on the stage of the evaluation cycle. Second, this finding also adds to the debate between low-fidelity and high-fidelity proponents in the context of UX design (see, Rudd et al., 1996), and suggests a contingent answer. Specifically, intuition, and common received wisdom from design literature (Tullis, 1990), suggests that low-fidelity tests that yield noisy measurements may be more valuable at the beginning stages, but high fidelity tests are possibly necessary before finalizing design choices.[7] In contrast, we find that the existing "intuition" cited above (i.e., use low fidelity early and high fidelity later) is not always true and there exists

[6] Note that it is easy to show that the firm's net profits are monotone increasing in τ_i. Hence, assuming that $w(\tau)$ is non-decreasing involves no loss in generality (since no measurement technology that is pricier **and** lower fidelity will ever be chosen by the firm).

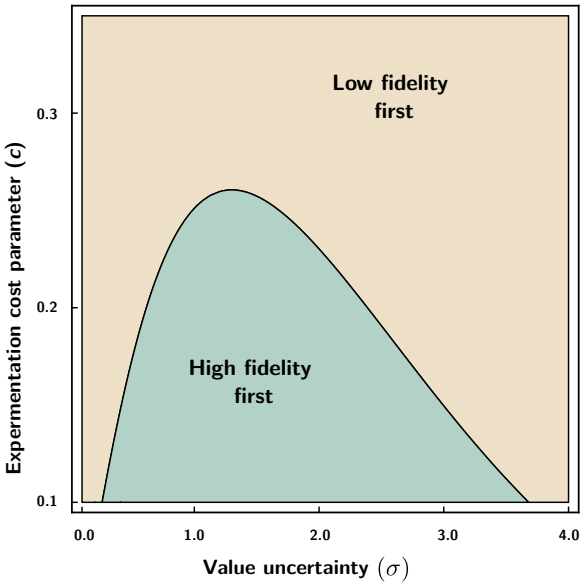
[7] We should note that this intuition might not be exactly comparable since it was developed in the context of testing a single alternative (and not testing multiple alternatives as in our setting).

conditions under which the firm finds it optimal to employ higher fidelity evaluations to start with, and then progressively use lower fidelity evaluations in the later stages of the evaluation cycle. More concretely, as characterized in the Proposition 4 and also shown in Figure 2, the optimal sequence of measurement technologies may (*Case A*) start off with low fidelity and then progressively use higher and higher fidelities, or (*Case B*) start off with high fidelity and then progressively use lower and lower fidelities.

To understand the intuition behind this surprising result, it is useful to consider how the signal provided by the evaluations interacts with ex-ante uncertainty of the alternatives (see figure 2). When the prototypes have very low variability (σ), then any measurement that is much different than the mean is possibly the result of measurement error. Consequently, the evaluations are most valuable to eliminate really terrible alternatives rather than explicitly choose one to be the best. Evaluations with low fidelity are reasonably effective in this exercise and so the firm benefits by saving on costs and starting out with low fidelity evaluations. As the evaluation cycle proceeds, using higher fidelity evaluations to select one starts becoming important, resulting in the firm finding it valuable to increase the fidelity of its evaluations.

In contrast, for moderate σ, a previously received low or high evaluation could be the result of prototype's intrinsic variability or because of noise in the evaluation, and identifying the source of deviation is important to determine whether the alternative should be selected or rejected. Consequently, by starting out with higher fidelity evaluations, we can identify the truly good prototypes earlier (thus avoiding unnecessary future tests), and so the optimal fidelity decreases over the test cycle. Still, for much higher prototype variability σ, a very high (or very low) measurement most likely came about because of the prototype's own intrinsic uncertainty, we would find less value in obtaining really high fidelity measurements. Consequently, we can start off with lower fidelity evaluations, and then progressively refine them if and as needed. In fact, if outcomes look favorable early on in the test cycle, then the purpose of later tests is to primarily gain confidence that there are no hidden high-potential solutions in unexplored parts of the design space. Consequently, lower fidelity tests, which benefit from lower costs, may be utilized later in the test cycle.

In contrast to the effect of prototype's intrinsic uncertainty explained above, the effect of relative cost of fidelity on the firm's choice is more straightforward. Intuitively, when the marginal cost for increased fidelity is relatively low (i.e., low c), the firm is indifferent, from the cost perspective, to using a high or low fidelity measurement. Hence, in this region, since a high-fidelity measurement improves the firm's revenues without imposing significant extra cost, it is optimal for the firm to use a higher fidelity evaluation to start off, and then follow it up with lower fidelity evaluation if needed. For higher c, the cost of increased measurement fidelity becomes more significant, and the firm maximizes their



Note: $w(\tau) = c\tau$ *and* $\tau \in \{0.2, 1\}$ *in the plotted example with two alternatives.*

Figure 2: Optimal sequencing of measurement fidelity

profits by minimizing the evaluation costs and selecting cheaper lower-fidelity evaluation first, and then following it up with higher fidelity evaluation (only when needed).

Proposition 4 assumed that the firm, ex-ante, can make a different choice of measurement technology for every prototype (i.e., τ_i is unique to i^{th} prototype). Still, in some situations, the firm might have considerably fewer degrees of freedom in this choice and might need to make the same choice of fidelity (independent of the prototype being evaluated). For instance, consider the case where the firm employs a specific pool of people to rate a user interface. In such situations, it appears more reasonable to assume that, once the firm has chosen the evaluation fidelity (by choosing the specific pool of testers), then every evaluation of any of their N designs will have identical evaluation fidelities $\tau_i = \tau$, and will impose identical costs $w_i = w$. With these assumptions, the next proposition characterizes the firm's expected revenues/cost/profit and their optimal choice of measurement technology.

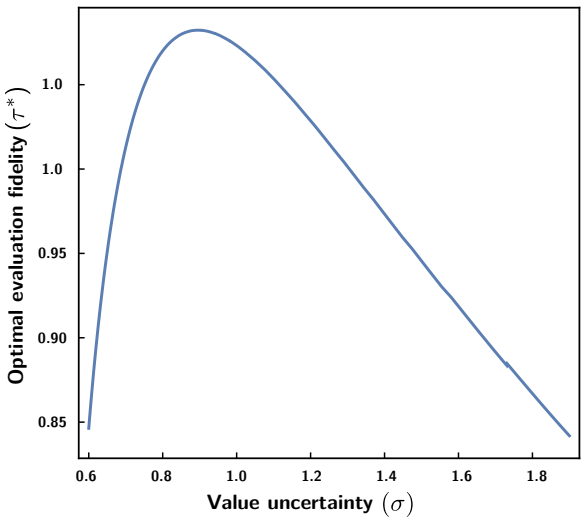
Proposition 5. *For an arbitrary* N*, the focal firm's expected revenue is given by*

$$\mu + {}^{\sigma}/_{\theta} R\left({}^{w\theta}/_{\sigma}\right) + \frac{w}{1 - \Phi\left(R\left({}^{w\theta}/_{\sigma}\right)\right)} + O\left(\exp\left(-N\right)\right)$$

and the expected cost by

$$\frac{w}{1 - \Phi\left(R\left({}^{w\theta}/_{\sigma}\right)\right)} + O\left(\exp\left(-N\right)\right)$$

When $N \to \infty$*, the measurement technology* τ^* *that maximizes the firm's profits is increasing in* σ *for low enough* σ*, and is decreasing in* σ *for high enough* σ*.*



Note: $w(\tau) = 0.1\tau$ *in the example.*

Figure 3: Optimal choice of fidelity as a function of prototype's intrinsic uncertainty

The proposition offers the revenues and costs for arbitrary N. Unsurprisingly, a complete closed-form characterization proves intractable (since even the simpler Weitzman, 1979 model, with perfect fidelity evaluations, does not possess one). Still, as the number of alternatives increases, the proposition offers a simple closed-form expression, where the error declines geometrically fast. From here onward, we shall explicitly assume that $N \to \infty$, for the sake of analytical tractability. Still, given that our subsequent analysis relies on this assumption, we shall also offer additional theoretical analysis in §5.2 demonstrating that our main qualitative results are retained even for small N.

Figure 3 illustrates the effect of prototype's intrinsic uncertainty on the focal firm's choice of optimal evaluation fidelity. Intuitively, higher fidelity measurements are less valuable when the evaluations are likely to fall much above or much below the ex-ante mean, since in these cases we would either immediately stop and accept, or reject and continue. Thus, when the alternatives have high uncertainty (i.e high σ), as the evaluations are more likely to fall further away from the mean, the evaluations are less valuable and the firm should save costs by choosing low fidelity measurements. In contrast, when the values are more likely to fall closer to the mean (i.e., moderate σ), the firm benefits from using a higher fidelity evaluation (as this allows them to more easily accept/reject these "boundary" cases). Still, when alternatives exhibit very low uncertainty (i.e., low σ), all evaluations are less informative in any case (as the most likely true value is in fact close to the mean), and the firm chooses cost-efficient low fidelity evaluations. Taken together, the optimal fidelity is increasing and then decreasing in the uncertainty of the alternatives.

5 Extensions

We explore three distinct extensions in this section: First, our main model assumed that while the firm's evaluation policy is dynamic (i.e., the firm can base their decisions on previously observed evaluations), the choice of measurement fidelity is static (i.e., needs to be made before commencing the evaluation cycle). §5.1 explores how optimal policy is affected when the fidelity can also be dynamically chosen. Second, given the analytical difficulty in finding closed-form results for intermediate N, our main model presented earlier could not speak to how the optimal fidelity is affected by the number of alternatives N. We address this in §5.2 by critically examining our assumption of $N \to \infty$.[8] Finally, we motivated the context of low fidelity evaluations by offering an example of outsourced evaluation using experimentation platforms. Still, our model and analysis did not consider the platform as an active participant. We remedy this in §5.3 by explicitly considering the platform's choices, and in specific their choice of which measurement technology to offer and how to price it.

5.1 Dynamic Choice of Measurement Fidelity

We assumed in §4.2 that the choice of measurement technology is static and decided before commencing the test cycle. Still, it is feasible that, depending on the context, the firm may be able to observe the results from the initial evaluation before deciding on the fidelity that they need from the next evaluation. We shall examine such dynamic choice of measurement technology with the goal of understanding if (and how) past evaluations influence future choice of fidelity.

As in our base-line model, at each stage, the firm chooses whether to stop and accept one of the previously tested prototypes. However, in contrast to the static setting, if the firm chooses not to terminate the evaluation cycle, then they may choose the prototype to evaluate next and can specify at what fidelity (τ) they require the evaluation. Given the complexity of this dynamic program, it is unsurprising that the optimal test schedule turns out to be intractable to characterize in the general case. Because of this, in the analysis that follows, we shall only examine a simplified version of the model presented in §4, and assume exactly two prototypes $N = 2$. For notational convenience, define $\Delta_1 = \mu_1 + (m_1 - \mu_1)\,\frac{\sigma_1^2}{\sigma_1^2+\tau_1^2} - \mu_2$ where m_1 is the measurement from the first evaluation.

Proposition 6. *Optimal test fidelity τ_2^* is increasing in Δ_1 when $\Delta_1 > 0$, and is decreasing in Δ_1 when $\Delta_1 < 0$.*

Intuitively, if the best value we have seen so far is very low, then the alternative that is examined next is highly likely to be selected, which reduces the value of extra information that a higher fidelity measurement yields. At the same time, if the best value we have seen so far is high (relative to the

[8] We are grateful to an anonymous reviewer and editor who pushed us to consider the small N setting.

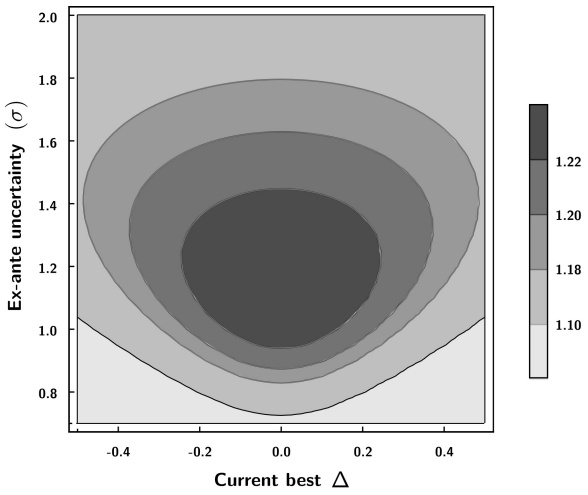


Figure 4: Choice of fidelity as a function of best previously observed evaluation and ex-ante uncertainty *Note:* $\sigma = 1, \mu = 0$ *and* $w(\tau) = 0.1\tau$ *in the example.*

expected value of new alternative), then, it is unlikely that the next examined alternative will be selected, once again reducing the value of additional information that a higher fidelity measurement yields. Thus, the value of higher fidelity measurement is greatest when the best that we have seen so far is neither terrible nor exceedingly good. Consequently, higher fidelity measurements of future alternatives are of value only when the measurements so far are modest. Finally, similar to our finding in Proposition 5, the optimal dynamically chosen fidelity is highest (least) when the intrinsic uncertainty is moderate (low or high). These results are graphically illustrated in Figure 4.

5.2 Dependence of optimal fidelity on the number of alternatives N

Proposition 5 holds for an arbitrary N, with the $O(\exp(-N))$ term decreasing exponentially fast with N. Relying on this fact, we stated and proved the firm's optimal fidelity choices (Proposition 5) for the case of large N. There are two distinct reasons why, we believe, this assumption is reasonable - first, from the technical perspective, the geometrically decreasing errors would suggest that the qualitative results are retained for reasonable sized N; and second, from the practical perspective, most firms terminate testing not because they ran out of alternatives to test, but because they found satisfactory performance (i.e., N is hardly, if ever, a binding constraint). Still, to "stress test" our results, consider the extreme case of $N = 1$. That is the firm has exactly one design that they can possibly evaluate. In this setting, if the firm chooses a measurement technology τ at cost $w(\tau)$, then the firm's expected payoff is given by $E\left[\max\left\{0, \mu + \mathbf{Z}\frac{\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}\right\}\right] - w(\tau)$. Thus, the equivalent of Proposition 5, for the case for $N = 1$ becomes

Proposition 5′ *The expected revenue is given by* $(\sigma/\theta)\left(\frac{\mu\theta}{\sigma}\Phi\left(\frac{\mu\theta}{\sigma}\right) + \frac{e^{-\frac{\mu^2\theta^2}{2\sigma^2}}}{\sqrt{2\pi}}\right)$*, and the expected cost by* $w(\tau)$*. For* $\mu = 0$*, the measurement technology* τ^* *that maximizes the firm's profits is increasing in*

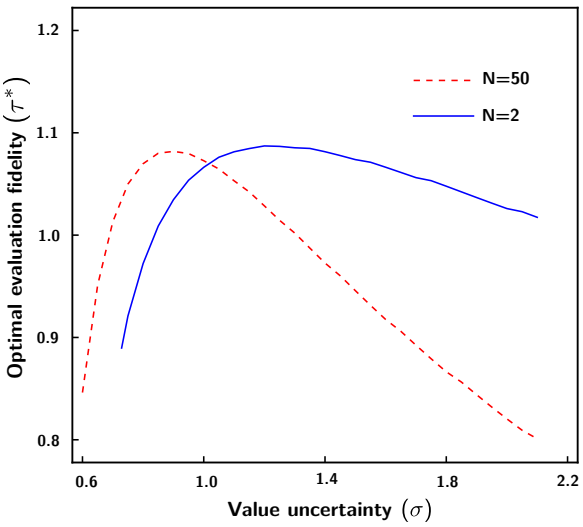


Figure 5: Optimal choice of fidelity as a function of prototype's intrinsic uncertainty
Note: $w(\tau) = 0.1\tau$ *in the example.*

σ for low enough σ, and is decreasing in σ for high enough σ.

Proposition 5′ show that the sensitivity results claimed in our main model (with $N \to \infty$) are true even in our "stress test" scenario of $N = 1$. In fact, these qualitative results continue to hold even for intermediate values of N (> 1), as graphically illustrated in Figure 5. As may be observed from this analysis, the claimed non-monotonicity for the firm's optimal choice of fidelity turns out to be robust, and it is when the uncertainty is moderate that the highest fidelity measurements are used.

Figure 5 also suggests an intriguing interaction between the number of available alternatives and the value uncertainty. Specifically, when the value uncertainty is low, it may be seen that having more alternatives (an increase from $N = 2$ to $N = 50$) increases the optimal evaluation fidelity τ^*. In other words, for low σ, evaluation fidelity and number of alternatives are (strategic) complements. Intuitively, when the value from the different alternatives are more predictable and possibly more similar to each other (low σ), the firm needs to test each alternative more accurately to distinguish between them. Consequently, they use higher fidelity evaluations with more alternatives.

In contrast, when value uncertainty is high, it can be seen that having more alternatives (an increase from $N = 2$ to $N = 50$) decreases the optimal evaluation fidelity τ^*. That is, for high σ, evaluation fidelity and number of alternatives are (strategic) substitutes. Intuitively, when value uncertainty is high, then the alternatives are more likely to be very different in value, making the job of distinguishing between them easier even with lower fidelity evaluations. Thus, when σ is high, the firm should use higher fidelity evaluations only when they have fewer alternatives.

While the results offered so far pointed to the robustness of our main result and suggested an interesting interaction between value uncertainty and the number of alternatives, it is inadequate

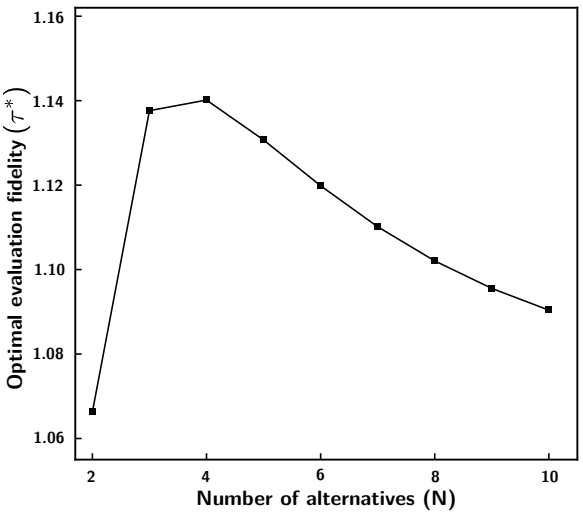


Figure 6: Optimal choice of fidelity as a function of total number of alternatives N
Note: $w(\tau) = 0.1\tau$ *and* $\sigma = 1$ *in the example.*

to understand if and how the optimal chosen fidelity depends on the total number of alternatives available to the firm. Unfortunately, finding analytical results, in this intermediate region of N, proves intractable (see proof of Proposition 4 for details). To partially address this, we conducted additional numerical experiments, an example of which is shown in Figure 6.

It is interesting to observe that the highest fidelity is chosen when the firm has a moderate number of alternatives to test. Intuitively, with very few alternatives, the role of evaluation is primarily to avoid selecting a negative valued alternative, and for this purpose a lower fidelity evaluation proves sufficient.[9] When the number of alternatives increases, then greater fidelity allows the firm to more accurately compare between alternatives and thus select the best one. Consequently, the firm benefits from pursuing higher fidelity evaluations. But when the number of alternatives is very large, then at least one of these alternatives is going to exhibit very high performance, and consequently be easy to identify even with lower fidelity evaluations. Thus, the firm saves on evaluation costs and employs lower fidelity evaluations when number of alternatives is very large.

5.3 Platform's problem: Optimal Design and Pricing of Tests

The model and analysis thus far considered the case of a firm that conducts all the stages of the product development, including the testing and evaluation phase, internally. However, as previously discussed, the rise of specialized experimentation platforms has led to a gradual shift toward outsourcing evaluation and testing, even as firms retain control over the rest of the design and prototyping decisions. These platforms allow product development firms to capitalize on scale and efficiency, resulting in a

[9] As an extreme example, consider the case of $N = 1$. In this case, the role of evaluation is only to rule out negative valued alternatives.

more streamlined and cost-effective development process. Still, the lack of visibility into the actual evaluation process and the possible ambiguity in the product development firm's requirements makes the problem of noisy evaluations even more salient. To understand the implications of engaging these experimentation platforms and to understand the platform's decision-making, we now extend our model to make the platform an active participant who can choose the measurement technologies and the associated fees.

We shall assume, in this section, that the experimentation platform offers a single measurement technology τ. Moreover, we shall assume that the cost to the experimentation platform from (undertaking) a single evaluation of any design is $c(\tau)$. As in §4.2, we assume that the alternatives have identical means and standard deviations (μ and σ respectively). Under this assumption, it is helpful to first analyze the profits of the "client" firm and the various components that contribute to it, as outlined in Proposition 5. Specifically, when the experimentation platform offers a single measurement technology τ for the fee $w(\tau)$, the firm's net expected profit is given by

$$\underbrace{\mu + {}^{\sigma}\!/_{\theta} R\left({}^{w\theta}\!/_{\sigma}\right) + \frac{w}{1-\Phi\left(R\left({}^{w\theta}\!/_{\sigma}\right)\right)}}_{\text{client's expected revenues}} - \underbrace{\frac{w}{1-\Phi\left(R\left({}^{w\theta}\!/_{\sigma}\right)\right)}}_{\text{client's expected evaluation cost}} = \mu + {}^{\sigma}\!/_{\theta} R\left({}^{w\theta}\!/_{\sigma}\right)$$

where $\theta = \sqrt{1 + {}^{1}\!/_{(\tau\sigma)^2}}$.

While making choices with respect to the evaluation fee and optimal fidelity, the platform anticipates the decision making process of the firm as outlined in §4, and then makes decisions in manner that maximizes its own profits. In this regard, there are a few key aspects that the platform needs to keep in mind. First, the fee paid by the client is exactly the revenue obtained by the experimentation platform. Hence, the expected revenue received by the experimentation platform, when they offer measurement technology τ at a fee w is given by $\frac{w}{1-\Phi(R({}^{w\theta}\!/_{\sigma}))}$.

Second, while determining the evaluation fee, it is important to ensure that ${}^{\sigma}\!/_{\theta} R\left({}^{w\theta}\!/_{\sigma}\right)$ is non-negative. If this condition is not true, it would be optimal for the client firm to not ask for evaluations from the platform and instead launch the prototype (with only internal evaluations) and obtain the expected value μ.

Finally, as the cost to the experimentation platform from (undertaking) a single evaluation is $c(\tau)$, we may net out the expected cost to obtain the platform's net profits when they offer measurement technology τ for the fee $w(\tau)$.

Putting these together, the expected profits of platform can be written as

$$\Pi\left(\tau, w\left(\tau\right)\right) = \begin{cases} \frac{w-c}{1-\Phi(R(w\theta/\sigma))} & \text{if } w\theta/\sigma \leq R^{-1}\left(0\right) \\ 0 & \text{o/w} \end{cases}$$

For any given fidelity of the measurement technology, the platform's profit maximizing evaluation fee should balance several objectives. Not surprisingly, number of evaluations requested by the client firm decreases in the fee charged by the platform (i.e., $\frac{1}{1-\Phi(R(w\theta/\sigma))}$ is decreasing in w), and the profit per evaluation is increasing in the fee (i.e., $w-c$ increases in the value of fee w). Thus, the effect of higher fees on the overall revenue from the client is ambiguous. Still, beyond a point ($w > \sigma R^{-1}(0)/\theta$), higher fees will induce the client to forgo evaluations (on the platform) and launch their product with only internal evaluations, thus reducing the platform's revenues.

The next proposition characterizes the optimal evaluation fee charged by the platform, and the client's induced experimentation behavior, for an arbitrary measurement technology τ.

Proposition 7. *For a given evaluation fidelity τ, the experimentation platform charges the profit maximizing evaluation fee $w^*\left(\tau\right) = \frac{R^{-1}(0)\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}$. Furthermore, $w^*\left(\tau\right)$ is increasing in σ and τ.*

The proposition shows that the optimal fee charged the platform, as we might intuitively expect, is greater/lesser when the client firm's alternatives exhibit greater/lesser uncertainty or when the evaluation fidelity is high/low. These two comparative statics results also have an interesting implication to the induced experimentation cycle followed by the client firm. Specifically, while greater measurement fidelity can increase the number of evaluations requested by the client, this same greater measurement fidelity also allows the platform to charge a higher fee, which will then decrease the number of requested evaluations. Thus, in our model of optimal pricing of evaluations, the expected number of evaluations requested by the client firm is always the same (and not increasing in the fidelity of the measurement technology).

The same insight holds for the case where the client firm's alternatives exhibit greater/lesser uncertainty (σ). Specifically, while higher uncertainty increases the client firm's incentives to undertake more evaluations, the experimentation platform, anticipating this, will charge a higher fee, which lowers the client's incentive to pursue more evaluations. Thus, the expected number of evaluations requested by the client firm is always the same and not increasing in the uncertainty of the alternative.

Having characterized the optimal fee $w^*\left(\tau\right)$ for an arbitrary measurement technology τ, we next characterize the profit maximizing measurement technology τ_p^* that will be offered by the experimentation platform.

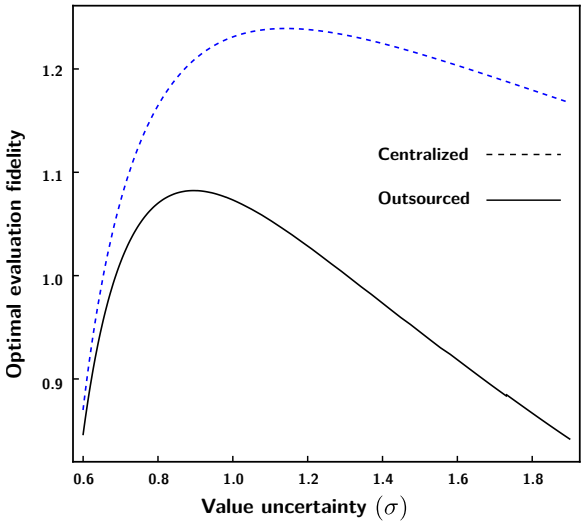


Figure 7: §Platform's optimal provisioning of measurement fidelity
Note: $c(\tau) = 0.1\tau$ *in the example.*

Proposition 8. *The profit maximizing measurement technology offered by the platform is*

$$\tau_p^* = \arg\max_{\tau} \left\{ \frac{R^{-1}(0)\,\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} - c(\tau) \right\}$$

Moreover, τ_p^* *is increasing in* σ *for sufficiently low* σ*, and decreasing in* σ *for sufficiently high* σ*.*

The solid curve in Figure 7 illustrates the effect of client prototype's intrinsic uncertainty on the experimentation platform's optimal provisioning of measurement fidelity. Interestingly, the intuition mirrors the one we developed for the case of client chosen measurement technology (shown in Proposition 5). Specifically, it is only when the client's prototypes have moderate intrinsic uncertainty that the platform will offer a high-fidelity evaluation. When client's prototypes exhibit very low uncertainty, any evaluation is less valuable and the client finds no real need for costly high-fidelity evaluations, and hence, the platform offers only low-fidelity evaluations. When client's prototype exhibits very high uncertainty, measurements are likely to fall much above or much below the mean, the former resulting in an obvious "stop and accept" decision and the latter in an obvious "reject and continue" decision. Thus, when the client's prototypes have high uncertainty (σ), the evaluations are less likely to result in "boundary" cases (where it is unclear whether to stop or to continue), and the client finds limited value in costly high fidelity evaluations. Consequently, the platform offers only a low-fidelity measurement in this case.

In addition to the characterization of the optimal evaluation fee for arbitrary evaluation fidelities (in Proposition 7), it is also of interest to consider the optimal fee corresponding to the optimal measurement technology (i.e., $w^*\left(\tau_p^*\right)$). Recall that $w^*(\cdot)$ is increasing in σ, and τ_p^* is decreasing in

σ when σ changes from moderate to high values. Taken together, this points to an intriguing result: when the client's intrinsic uncertainty increases from moderate to high values, the platform offers lower fidelity evaluations, but at a higher price.

Related to the question of optimal platform offered fidelity is the question of which measurement fidelity would jointly maximize the client and experimental platform's total payoffs. Note that the joint profits are given by $\frac{w-c}{1-\Phi(R(w\theta/\sigma))}+\mu+\sigma/\theta R\left(w\theta/\sigma\right)$. Obviously, this profit is maximized when $w=c$, since this avoids double marginalization (i.e., platform sets their fee at cost). Thus, from the perspective of a centralized total welfare maximizing decision-maker, the efficient fidelity is

$$\tau_c^* \quad = \quad \arg\max_{\tau} \left\{\sigma/\theta R\left(c\theta/\sigma\right)\right\}$$

Figure 7 (dashed blue curve) shows the efficient fidelity τ_c^*. It may be noted that, consistent with the intuition developed in Proposition 5 and 8, the efficient fidelity is increasing and then decreasing with client prototypes' intrinsic uncertainty σ. Moreover, Figure 7 also shows that the efficient fidelity τ_c^* is always higher than the decentralized fidelity τ_p^*. Intuitively, the inability to appropriate the full economic rents cause the platform to offer less than efficient evaluations.

6 Discussion & Conclusions

Prototyping and testing are critical facets of the product development process, playing a fundamental role in helping firms manage the inherent uncertainties associated with creating new products. Recent developments in rapid prototyping technologies, including technologies that enable cheaper low fidelity tests, have introduced new tradeoffs that balance evaluation precision against the prototyping costs.

We develop a model of "noisy sequential search" where the focal firm sequentially obtains evaluation of its prototypes, and where these evaluations are a noisy signal of the prototype's "true value." In this setting, we show that despite the complexity of the underlying context, the optimal dynamic policy is a threshold reservation price policy; i.e., the firm would be able to make its decision on whether to continue or stop testing solely based on how the measurement compares to an ex-ante computed threshold. Interestingly, despite the simplicity of the optimal policy, its operationalization leads to some surprising results. First, when the evaluations are noisy, it might sometimes be optimal for the firm to select and launch a prototype that did not yield the highest measurement. This is in contrast to perfect fidelity evaluations, where a firm should always select the prototype with the highest evaluation. Second, this effect becomes more pronounced as the fidelity in the evaluations decreases. This suggests that firms using low fidelity evaluations for testing should adjust their selection criteria for prototypes.

Given the nuanced relationship between the fidelity of evaluations and a firm's selection criteria, we

		Ex-ante Uncertainty		
		Low	Med	High
Cost of Evaluations	High	Low Fidelity Minimal Testing	Modest Fidelity Limited Testing	Low Fidelity Moderate Testing
	Low	Modest Fidelity Minimal Testing	High Fidelity Moderate Testing	Modest Fidelity Extensive Testing

Table 1: Framework for "noisy" experimentation

also examine how a firm should choose evaluations that balance the cost and fidelity of evaluations. Our results demonstrate that the optimal choice of evaluation fidelity has a non-monotonic relationship with prototypes' ex-ante uncertainty. Specifically, the firm should seek to obtain high fidelity evaluations only when the ex-ante uncertainty of the alternatives is moderate. Under these conditions, the firm faces substantial uncertainty about when they should continue the evaluation process or terminate it, and the higher fidelity evaluations provide the necessary guidance to make informed decisions.

The tradeoff between higher fidelity and higher cost, and how the uncertainty of the design alternatives determines the optimal choice, extend even to cases when the focal firm might have the capability to dynamically change the chosen fidelities as the test cycle proceeds. Specifically, we find that the firm should choose high fidelity only when the past measurements are neither much better nor much worse than expected. We extend our analysis to situations where the number of available design alternatives is more limited. As very large number of available alternatives guarantees that at least one is exceptional and will stand out (irrespective of the evaluation fidelity), we find that it is with moderate number of alternatives that the firm employs the highest fidelity evaluation.

Finally, we extend our analysis to settings where the focal firm engages an outside experimentation platform to evaluate the prototypes. Even in this "outsourced" setting, we find an interesting alignment between the two in terms of when maximum fidelity should be offered. Specifically, similar to the case with firm's choice of optimal fidelity, the evaluation fidelity offered by the platform is also increasing and then decreasing in the uncertainty of the focal firm's prototypes. Still, since the focal firm has greater upside potential with increasing ex-ante uncertainty, the platform charges a higher fee for their evaluations when the client's prototypes have greater uncertainty. Taken together, these provide an interesting observation: when the focal firm's ex-ante uncertainty increases from moderate to high values, the platform should offer lower fidelity evaluations, but at a higher price.

Our results allow us to draw a managerially relevant framework of noisy experimentation illustrated in Table 1. Firstly, noisy experimentation depends on the cost of evaluations (a measurement specific

attribute), with greater/lesser fidelity associated with lower/higher evaluation costs. The firm's ex-ante uncertainty (a prototype specific attribute), however, has a more subtle effect on noisy experimentation. The value of fidelity is highest when the ex-ante uncertainty is moderate and additional evaluations are most valuable when the ex-ante uncertainty is highest.

Taken together, when the cost of evaluation and the ex-ante uncertainty is low (as in the UX design mentioned in Section 1), the firm should plan for shortened evaluation cycles at modest evaluation fidelity. In contrast, when the cost of evaluation and the ex-ante uncertainty is higher (as in the smart chatbot mentioned in Section 1), the firm plans for longer evaluation cycles at modest fidelity. Interestingly, high fidelity evaluations are unnecessary in either of these settings. It is when cost of evaluation is low and ex-ante uncertainty is moderate (which is plausibly the case in feature management/enhancements when a firm is testing out different features for their websites and apps[10]), that the focal firm should plan on longer evaluation cycles at high fidelity.

The current article, to the best of our knowledge, offers one of the few complete generalization of the classic Weitzman (1979) model to account for less than perfect measurements. But in using the model to study the context of noisy experimentation, we necessarily had to make several simplifying assumptions. For the case of experimentation platforms and outsourced evaluations, we abstracted away from agency issues for the most part and assumed that the experimentation platform will only evaluate and accurately report the performance of a requested prototype. These assumptions, while reasonable in our context, might be less acceptable in at least two related settings - first, cases where the experimentation platform's measurement technology is a black-box (and hence, unverifiable), and second, when the experimentation platform has significant domain-specific design competencies. In the former, the platform may shirk from costly evaluations, and thus might necessitate more complex contracts to obtain alignment between the two parties (see for instance, Schlapp and Schumacher, 2022). In the latter, the focal firm firm might be able to outsource the overall dev-test cycle (see for instance, Zorc et al., 2023).

We also assumed that the focal firm does not undertake evaluations of multiple alternatives in parallel, and instead engages in sequential evaluations of the alternatives. Still, to the extent that the time required for sequential evaluations imposes additional costs, the firm might benefit from parallel evaluations. While a full exploration of such settings and a comparison between sequential and parallel evaluations lie beyond the scope of current article, preliminary analysis that extends our baseline model to parallel tests seem to suggest that our results linking optimal fidelity to the ex-ante value uncertainty of the alternative may survive.

Finally, our model assumes that a given alternative is tested (perfectly or imperfectly) exactly once.

[10] See Optimizely.com for an instance of such "feature management" offerings.

While this is true in situations where the test options are more limited, in cases where the firm has multiple test options available for each alternative, one may need to consider models which incorporate the possibility of repeated tests of the same alternative (like multi-armed bandit models). Extending the current model in these directions to capture such richer space of possibilities remain tasks for future research.

References

Adam, K. 2001. Learning while searching for the best alternative. *Journal of Economic Theory* **101**(1) 252–280.

Bansal, S., G. J. Gutierrez, J. R. Keiser. 2017. Using experts' noisy quantile judgments to quantify risks: Theory and application to agribusiness. *Operations Research* **65**(5) 1115–1130.

Berman, R., C. Van den Bulte. 2022. False discovery in a/b testing. *Management Science* **68**(9) 6762–6782.

Bhaskaran, S. R., V. Krishnan. 2009. Effort, revenue, and cost sharing mechanisms for collaborative new product development. *Management Science* **55**(7) 1152–1169.

Bhat, N., V. F. Farias, C. C. Moallemi, D. Sinha. 2020. Near-optimal ab testing. *Management Science* **66**(10) 4477–4495.

Busche, L. 2014. The skeptic's guide to low-fidelity prototyping. *Smashing magazine* .

Crama, P., B. De Reyck, N. Taneri. 2017. Licensing contracts: Control rights, options, and timing. *Management Science* **63**(4) 1131–1149.

Dahan, E., H. Mendelson. 2001. An extreme-value model of concept testing. *Management science* **47**(1) 102–116.

Doval, L. 2018. Whether or not to open pandora's box. *Journal of Economic Theory* **175** 127–158.

Erat, S., S. Kavadias. 2008. Sequential testing of product designs: Implications for learning. *Management Science* **54**(5) 956–968.

Gerardi, D., L. Maestri. 2012. A principal–agent model of sequential testing. *Theoretical Economics* **7**(3) 425–463.

Jacobs, P. F. 1992. *Rapid prototyping & manufacturing: fundamentals of stereolithography*. Society of Manufacturing Engineers.

Joyner, J. S., A. Kong, J. Angelo, W. He, M. Vaughn-Cooke. 2022. Development of low-fidelity virtual replicas of products for usability testing. *Applied Sciences* **12**(14). doi:10.3390/app12146937. URL `https://www.mdpi.com/2076-3417/12/14/6937`.

Kettunen, J., A. Salo. 2017. Estimation of downside risks in project portfolio selection. *Production and Operations Management* **26**(10) 1839–1853.

Koning, R., S. Hasan, A. Chatterji. 2022. Experimentation and start-up performance: Evidence from a/b testing. *Management Science* **68**(9) 6434–6453.

Kornish, L. J., J. Hutchison-Krupat. 2017. Research on idea generation and selection: Implications for management of technology. *Production and Operations Management* **26**(4) 633–651.

Loch, C. H., C. Terwiesch, S. Thomke. 2001. Parallel and sequential testing of design alternatives. *Management Science* **47**(5) 663–678.

McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management science* **31**(11) 1372–1389.

Özkan-Seely, G. F., D. C. Hall, J. Hutchison-Krupat. 2023. Search for the best alternative: An experimental approach. *Decision Sciences* .

Pham, D. T., R. S. Gault. 1998. A comparison of rapid prototyping technologies. *International Journal of machine tools and manufacture* **38**(10-11) 1257–1287.

Rahmani, M., K. Ramachandran. 2021. Delegating innovation projects with deadline: Committed vs. flexible stopping. *Management Science* **67**(10) 6317–6332.

Rahmani, M., G. Roels, U. S. Karmarkar. 2017. Collaborative work dynamics in projects with co-production. *Production and Operations Management* **26**(4) 686–703.

Roels, G., U. S. Karmarkar, S. Carr. 2010. Contracting for collaborative services. *Management Science* **56**(5) 849–863.

Rudd, J., K. Stern, S. Isensee. 1996. Low vs. high-fidelity prototyping debate. *interactions* **3**(1) 76–85.

Schlapp, J., G. Schumacher. 2022. Delegated concept testing in new product development. *Operations Research* **70**(5) 2732–2748.

Smith, J. E., R. L. Winkler. 2006. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science* **52**(3) 311–322.

Sommer, S. C., E. Bendoly, S. Kavadias. 2020. How do you search for the best alternative? experimental evidence on search strategies to solve complex problems. *Management Science* **66**(3) 1395–1420.

Terwiesch, C., Y. Xu. 2008. Innovation contests, open innovation, and multiagent problem solving. *Management science* **54**(9) 1529–1543.

Thomke, S. H. 1998. Managing experimentation in the design of new products. *Management Science* **44**(6) 743–762.

Thomke, S. H. 2003. *Experimentation matters: unlocking the potential of new technologies for innovation*. Harvard Business Press.

Tian, Z., H. Gurnani, Y. Xu. 2021. Collaboration in development of new drugs. *Production and Operations Management* **30**(11) 3943–3966.

Tullis, T. 1990. High-fidelity prototyping throughout the design process. *Proceedings of the Human Factors Society 34th Annual Meeting (Santa Monica, CA, Human Factors Society 1990)*. 266.

Weitzman, M. L. 1979. Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society* 641–654.

Zorc, S., I. Tsetlin, S. Hasija, S. E. Chick. 2023. The when and how of delegated search. *Operations Research* .

Technical Appendix: Optimal Prototyping with Noisy Measurements

Proof of Theorem 1

For ease of exposition, refer to the main search as problem $\mathcal{P}$.

Claim 1. Consider a new search problem $\mathcal{P}_2$ (without any evaluation uncertainty) where the prototypes i $(= 1, \cdots, N)$ is distributed as a random variable X_i with pdf $\int_{-\infty}^{\infty} f_i(P, X_i)\, dP$, and the cost of evaluation alternative i is w_i, and the value upon stopping and selecting prototype i is $Q_i(X_i)$ where

$$Q_i(m) \overset{\text{def}}{=} \int_{-\infty}^{\infty} \frac{f_i(P,m)}{\int_{-\infty}^{\infty} f_i(P,m)\, dm} P dP \tag{A.1}$$

This new search problem is equivalent to the original search problem $\mathcal{P}$.

Proof. In $\mathcal{P}$, for any given m_i, the expected value a searcher obtains upon stopping and choosing i is $Q_i(m_i)$, since the expectation of the true value of prototype i, conditional on observing a measurement m, is

$$Q_i(m) = \int_{-\infty}^{\infty} \frac{f_i(P,m)}{\int_{-\infty}^{\infty} f_i(P,m)\, dm} P dP$$

Moreover, in $\mathcal{P}$, the ex-ante distribution (pdf) of measurements $\mathbf{m_i}$ that we observe upon testing prototype i is given by $\int_{-\infty}^{\infty} f_i(P, m_i)\, dP$.

In the search problem $\mathcal{P}_2$, for any given m_i, the value a searcher obtains upon stopping and choosing i is $Q_i(m_i)$ (by definition). Moreover, m_i has the same distribution in problem $\mathcal{P}_2$ as in problem $\mathcal{P}$ (again by definition). Hence, we may set up a coupling between the two problems such that they have identical optimal policy and yield identical optimal values. ■

To simplify the exposition, assume that $Q_i(\cdot)$ is an increasing function with range $(-\infty, \infty)$.[11]

Claim 2. Consider a traditional Pandora's box search problem $\hat{\mathcal{P}}$ (i.e., one without any evaluation uncertainty) defined as follows: Alternative i $(i = 1, \cdots, N)$ has search cost w_i and value distribution

[11] The assumption makes it easier to explicitly write out the distribution of value $Q_i(X_i)$ from selecting alternative i in search problem $\mathcal{P}_2$. Still, it is to be noted that the assumption is merely meant for notational simplicity, and even for arbitrary functions, we can easily write out the distribution of $Q_i(X_i)$.

Also, note that the assumption states a fairly intuitive property that any reasonable measurement should probably possess, namely that the expected value is increasing in the signal (measurement).

given by the pdf c

$$g_i(x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_i\left(P, Q^{-1}(x)\right) dP$$

The pandora's search problem $\hat{\mathcal{P}}$ is equivalent to the original (noisy measurements) search problem $\mathcal{P}$.

Proof. From Claim 1, we know that the equivalent search problem $\mathcal{P}_2$ has alternatives distributed as $\int_{-\infty}^{\infty} f_i(P, X_i)\, dP$ and where stopping and choosing i gives value $Q_i(X_i)$. Given the assumption that $Q_i(\cdot)$ is increasing, the ex-ante distribution of value upon stopping and choosing i is given by $\int_{-\infty}^{\infty} f_i\left(P, Q^{-1}(x)\right) dP$. Thus, the problem $\mathcal{P}_2$ is equivalent to Pandora's box problem $\hat{\mathcal{P}}$ where the value of alternative i is distributed as $g_i(x)$ defined above. ■

Claims 1 and 2 allow us to now use standard results from previous literature (Weitzman, 1979) to fully characterize the optimal policy for our problem with evaluation uncertainty. We give it below for completeness.

Let R_i be defined as the (smallest) solution of the equation $E[\max\{R, X_i\}] - R = w_i$ where $X_i \sim g_i(x)$.[12] Reorder the prototypes such that $R_i \geq R_j$ for $i < j$. For ease of exposition, let prototype 0 represent the option of not selecting any of the prototypes. Moreover, define V_i (which represents the expected value from the "best" prototype up to the i^{th}one) as follows

$$\begin{aligned} V_0 &\stackrel{\text{def}}{=} 0 \\ V_i &\stackrel{\text{def}}{=} \max\{V_{i-1}, Q_i(m_i)\} \text{ for } .i = 1, \cdots, N \end{aligned}$$

and let the corresponding "best" prototype be B_i. That is

$$\begin{aligned} B_0 &\stackrel{\text{def}}{=} 0 \\ B_i &\stackrel{\text{def}}{=} \begin{cases} B_{i-1} & \text{if } V_{i-1} > Q_i(m_i) \\ i & \text{o/w} \end{cases} \text{ for } .i = 1, \cdots, N \end{aligned}$$

■

[12] Such a solution exists since $E[\max\{R_i, Y_i\}] - R_i = E[\max\{0, Y_i - R_i\}]$ is non-increasing in R_i, and has the range $(0, \infty)$.

Proof of Theorem 2

The equivalent pandora's box problem has distribution $\mu_i + \left(\frac{\sigma_i^2}{\sqrt{\sigma_i^2+(1/\tau_i)^2}}\right)\mathbf{Z_i}$. Hence, the reservation price of this is the solution of $x = E\left[\max\left\{x, \mu_i + \left(\frac{\sigma_i^2}{\sqrt{\sigma_i^2+(1/\tau_i)^2}}\right)\mathbf{Z_i}\right\}\right] - w_i$. Rearranging the terms, we obtain that the reservation price is the solution of the following equation:

$$\left(\frac{\sqrt{\sigma_i^2+(1/\tau_i)^2}}{\sigma_i^2}\right)x - \mu_i \quad = \quad E\left[\max\left\{\left(\frac{\sqrt{\sigma_i^2+(1/\tau_i)^2}}{\sigma_i^2}\right)x - \mu_i, \mathbf{Z_i}\right\}\right] - \left(\frac{\sqrt{\sigma_i^2+(1/\tau_i)^2}}{\sigma_i^2}\right)w_i.$$

By our definition, $R(w)$ is the solution of $x = E[\max\{x, Z\}] - w$. Hence, we have that:

$$\frac{\sqrt{\sigma_i^2+(1/\tau_i)^2}}{\sigma_i^2}R_i - \mu_i \quad = \quad R\left(\frac{\sqrt{\sigma_i^2+(1/\tau_i)^2}}{\sigma_i^2}w_i\right),$$

which implies that

$$R_i \quad = \quad \mu_i + \frac{\sigma_i}{\sqrt{1+1/(\tau_i\sigma_i)^2}}R\left(w_i\frac{\sqrt{1+1/(\tau_i\sigma_i)^2}}{\sigma_i}\right).$$

To complete the proof, we note that the

$$\begin{aligned} p_i &= \mu_i + (m_i - \mu_i)\frac{\sigma_i^2}{\sigma_i^2 + 1/(\tau_i)^2} \\ &= \mu_i + (m_i - \mu_i)\frac{1}{1+1/(\tau_i\sigma_i)^2} \end{aligned}$$

is the expected value from choosing alternative i after observing measurement m_i. Moreover, equivalent pandora's box has the optimal policy where search proceeds onto the next highest reservation price, and terminating when the value p_i is above the next available reservation price. ■

Proof of Proposition 1

The alternative that is picked has a value V_i where

$$V_i \quad = \quad \arg\max_i\left\{\mu_i + (m_i - \mu_i)\frac{\sigma^2}{\sigma^2 + 1/\tau^2}\right\}.$$

Note that in general, this is not equal to $\arg\max_i\{m_i\}$.

Also note that the measurements are distributed as $m_i = \mu_i + \mathbf{Z}_i\sqrt{\sigma^2 + 1/\tau^2}$, and the expected true values are distributed as $p_i = \mu_i + \mathbf{Z}_i\frac{\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}$. Hence, the probability that $m_s > m_r$ when $p_r > p_s$ is

given by

$$\begin{aligned}\Psi(\tau) &= \Pr(m_s > m_r, p_r > p_s) \\ &\quad \Pr\left(\mu_r + \mathbf{Z}_r \frac{\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} > \mu_s + \mathbf{Z}_s \frac{\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}}, \mu_r + \mathbf{Z}_r \sqrt{\sigma^2 + 1/\tau^2} < \mu_s + \mathbf{Z}_s \sqrt{\sigma^2 + 1/\tau^2}\right) \\ &= \Pr\left((\mu_r - \mu_s) \frac{1}{\sqrt{\sigma^2 + 1/\tau^2}} < (\mathbf{Z}_s - \mathbf{Z}_r) < (\mu_r - \mu_s) \frac{\sqrt{\sigma^2 + 1/\tau^2}}{\sigma^2}\right)\end{aligned}$$

Noting that $\mathbf{Z}_s - \mathbf{Z}_r \sim \sqrt{2}\mathbf{Z}$ completes the calculation.

Finally, to show the second part of the proposition, note that when $\mu_r - \mu_s > 0$, then $(\mu_r - \mu_s) \frac{1}{\sqrt{\sigma^2 + 1/\tau^2}}$ is increasing in τ and $(\mu_r - \mu_s) \frac{\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$ is decreasing in τ. Hence, the probability $\Psi(\tau)$ is decreasing in τ. ■

Proof of Proposition 2

Note that $R_i = \mu + \frac{1}{\theta_i}[R(w_i\theta_i)]$ where $\theta_i = \frac{\sqrt{1 + 1/(\tau_i\sigma_i)^2}}{\sigma_i}$. It is easy to verify that the R_i is monotonically decreasing in w_i, since $R(\cdot)$ is a monotonic decreasing function. We can show that R_i is increasing in τ_i as follows: Obviously, θ_i is decreasing in τ_i. Hence, $\frac{d}{d\theta_i}(R_i) < 0$ will imply that $\frac{d}{d\tau_i}(R_i) > 0$. To prove that $\frac{d}{d\theta_i}(R_i) < 0$, we shall simply take the derivative to obtain

$$\begin{aligned}\frac{d}{d\theta_i}(R_i) &= \frac{1}{\theta_i} R'(w_i\theta_i) c_i - \frac{1}{\theta_i^2} R(w_i\theta_i) \\ &= \frac{w_i\theta_i R'(w_i\theta_i) - R(w_i\theta_i)}{\theta_i^2}\end{aligned}$$

Since $xR'(x) - R(x) < 0$ for all x, we have that $\frac{w_i\theta_i R'(w_i\theta_i) - R(w_i\theta_i)}{\theta_i^2} < 0$, and consequently, $\frac{d}{d\theta_i}(R_i) < 0$. Thus, it is proved that the reservation price R_i increases monotonically with τ_i.

Identical arguments show that the reservation price is increasing in σ (specifically, that θ_i is decreasing in σ_i). ■

Proof of Proposition 3

Consider the effect of τ_i on the expected outcomes under the optimal policy. We will test k or more alternatives if and only $p_0 \leq R_1$, and then $p_0 \leq R_2$ and $p_1 \leq R_2$, and then $p_0 \leq R_3$ and $p_1 \leq R_3$ and $p_2 \leq R_3$, and so on. As R_i are ordered, this implies that we will test k or more alternatives if and only if $p_0 \leq R_k$, $p_1 \leq R_k$, $\cdots$, $p_{k-1} \leq R_k$.

Since p_j is distributed as $\mu_j + \frac{\sigma_j^2}{\sqrt{\sigma_j^2 + (1/\tau_j)^2}} \mathbf{Z_j}$, we will test k (or more) alternatives if and only

$p_0 \leq R_k$ and $\mu_j + \frac{\sigma_j^2}{\sqrt{\sigma_j^2+\left(1/\tau_j\right)^2}}\mathbf{Z_j} \leq R_k$ for $j \in \{1, k-1\}$.

Thus, the probability of testing k (or more alternatives) is

$$\begin{aligned} P_k &= 1_{\{p_0 \leq R_k\}} \Pi_{j=1}^{k-1} \Pr\left(\mathbf{Z_j} \leq \frac{\sqrt{\sigma_j^2 + \left(1/\tau_j\right)^2}}{\sigma_j^2}\left(R_k - \mu_j\right)\right) \\ &= 1_{\{p_0 \leq R_2\}} \Pi_{j=1}^{k-1} \Phi\left(\frac{\sqrt{\sigma_j^2 + \left(1/\tau_j\right)^2}}{\sigma_j^2}\left(R_k - \mu_j\right)\right) \end{aligned}$$

where $\Phi(\cdot)$ is the cdf of the standard normal.

Since $\Phi(\cdot)$ is an increasing function (being a cdf), P_k is decreasing in τ_i if the reservation price $R_k > \mu_i$, whereas the probability of stopping after the first test is increasing in τ_i if the second highest reservation price $R_k < \mu_i$.

Note that the expected number of evaluations is given by $P_1 + P_2 + P_3 + \cdots$. Moreover, when τ_i decreases, P_j is unchanged for $j < i$; $P_i = 1$ for τ_i greater than a threshold and $= 0$ otherwise (since above the threshold $R_i > p_0$, which immediate eliminates the alternative); and for $j > i$, P_j increases or decreases as shown above.

Consequently, the expected number of evaluations should decrease in τ_i when μ_i is low enough, and should increase in the measurement fidelity τ_i when μ_i is high enough. ■

Proof of Proposition 4

For a given τ_i, optimal policy requires that the corresponding reservation prices are non-increasing. Since $R_i = \mu + \frac{\sigma}{\theta_i} R\left(w_i \frac{\theta_i}{\sigma}\right)$ where $\theta_i = \sqrt{1 + \left(1/\tau_i \sigma\right)^2}$ and $w_i = w(\tau_i)$, this implies that:

$$\frac{\sigma}{\theta_i} R\left(w_i \frac{\theta_i}{\sigma}\right) \geq \frac{\sigma}{\theta_j} R\left(w_j \frac{\theta_j}{\sigma}\right) \quad \forall \quad i < j$$

If $\frac{R(\theta w)}{\theta}$ is an increasing function of τ, we have that $\tau_1 > \tau_2 > \cdots \tau_n$. Similarly, for the case where $\frac{R(\theta w)}{\theta}$ is decreasing in τ, we will have that $\tau_1 < \tau_2 < \cdots \tau_n$.

Next, to explain the Figure 2, we need to demonstrate that $\frac{R(\theta w)}{\theta}$ is increasing in τ for high and low σ; and possibly decreasing only when σ is moderate. Note that the evaluation cost is assumed to be an increasing function in τ such that $w(\tau) = c\tau^{\alpha}$.

When σ is large, θ_i is nearly constant and decreases only slightly with τ. However, w increases with τ, and hence $R(\theta w)$ decreases with τ. So for large value of σ, we have that $\frac{R(\theta w)}{\theta}$ is decreasing in τ, and consequently $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$.

Now consider the case of really small values of σ. In this case, we have that $\theta \approx \frac{1}{\sigma\tau}$. Hence, $\frac{R(\theta w)}{\theta} \approx \sigma\tau R\left(\frac{c\tau^{\alpha-1}}{\sigma}\right)$. Taking the derivative w.r.t τ, we obtain

$$\begin{aligned}\frac{d}{d\tau}\sigma\tau R\left(\frac{c\tau^{\alpha-1}}{\sigma}\right) &\approx \sigma R\left(\frac{c\tau^{\alpha-1}}{\sigma}\right) + (\alpha-1)\tau R'\left(\frac{c\tau^{\alpha-1}}{\sigma}\right)c\tau^{\alpha-2} \\ & \quad (\alpha-1)\tau R'\left(\frac{c\tau^{\alpha-1}}{\sigma}\right)c\tau^{\alpha-2} < 0\end{aligned}$$

Thus, for very small σ, we again obtain that $\frac{R(\theta w)}{\theta}$ is decreasing in τ, and consequently $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$.

Observing the derivative above, we can see that it is only for moderate values of σ that the first term will be sufficiently positive so as to make the derivative positive, for which case we will have $\tau_1 \geq \tau_2 \geq \cdots \geq \tau_n$. Even in this case, the region may fail to exist (as may be noted for higher evaluation costs in Figure 2) ■

Proof of Proposition 5

Recall that, given our assumption of identical means $\mu_i = \mu$, standard deviations $\sigma_i = \sigma$, fidelity $\tau_i = \tau$, and cost $w_i = w$, the optimal policy is to continue testing as long as the measurement m is lower than the threshold M. Moreover, contingent on stopping, the expected value is $\mu + (m-\mu)\frac{\sigma^2}{\sigma^2+1/\tau^2} = \mu + (m-\mu)\frac{1}{\theta^2}$. Finally, recall that $m \sim N\left(\mu, \sigma^2 + 1/\tau^2\right) = N\left(\mu, \sigma^2\theta^2\right)$.

One may stop under two conditions - (i) "voluntarily" because the measurement exceeded the threshold, or (ii) "forced" because we ran out of alternatives to evaluate.

For the former case, the expected value upon stopping is

$$E\left[\mu + (\mathbf{m}-\mu)\frac{1}{\theta^2}|\mathbf{m} > \mu + \sigma\theta R\left(w\frac{\theta}{\sigma}\right)\right]$$

Some algebra verifies that this is equal to

$$\begin{aligned}& \mu + \frac{\sigma}{\theta}E\left[\frac{(\mathbf{m}-\mu)}{\sigma\theta}|\frac{\mathbf{m}-\mu}{\sigma\theta} > R\left(w\frac{\theta}{\sigma}\right)\right] \\ = \;& \mu + \frac{\sigma}{\theta}E\left[\mathbf{Z}|\mathbf{Z} > R\left(w\frac{\theta}{\sigma}\right)\right]\end{aligned}$$

Recall that $E\left[\mathbf{Z}|\mathbf{Z} > R\right](1-\Phi(R)) + R\Phi(R) = E\left[\max\left\{R, \mathbf{Z}\right\}\right] = R + w$. Hence, $E\left[\mathbf{Z}|\mathbf{Z} > R\right] =$

$R + \frac{w}{1-\Phi(R)}$. Combining, we obtain that the expected value upon voluntary stopping is given by

$$\begin{aligned} &= \mu + \frac{\sigma}{\theta}\left(R\left(w\frac{\theta}{\sigma}\right) + \frac{w\frac{\theta}{\sigma}}{1-\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}\right) \\ &= \mu + \left(\frac{\sigma}{\theta}R\left(w\theta/\sigma\right) + \frac{w}{1-\Phi\left(R\left(w\theta/\sigma\right)\right)}\right) \end{aligned}$$

Now consider the case of the expected value upon forced stopping. This situation occurs only when every single one of the tested alternatives measured less than the threshold. Consequently, the expected value can be rewritten as

$$\begin{aligned} &= E\left[\max_{i=1,\cdots,N}\left\{\mu + (\mathbf{m_i} - \mu)\frac{1}{\theta^2}\right\} \mid \max_{i=1,\cdots,N}\{\mathbf{m_i}\} < \mu + \sigma\theta R\left(w\frac{\theta}{\sigma}\right)\right] \\ &\quad \mu + \frac{\sigma}{\theta}E\left[\max_{i=1,\cdots,N}\left\{\frac{(\mathbf{m_i} - \mu)}{\sigma\theta}\right\} \mid \max_{i=1,\cdots,N}\left\{\frac{\mathbf{m_i} - \mu}{\sigma\theta}\right\} < R\left(w\frac{\theta}{\sigma}\right)\right] \\ &= \mu + \frac{\sigma}{\theta}E\left[\max_{i=1,\cdots,N}\{\mathbf{Z_i}\} \mid \max_{i=1,\cdots,N}\{\mathbf{Z_i}\} < R\left(w\frac{\theta}{\sigma}\right)\right] \\ &> \mu + \frac{\sigma}{\theta}E\left[\mathbf{Z_1} | \mathbf{Z_1} < R\left(w\frac{\theta}{\sigma}\right)\right] \\ &= \mu + \frac{\sigma}{\theta}R\left(w\frac{\theta}{\sigma}\right) - \frac{\sigma}{\theta}\left(\frac{w\frac{\theta}{\sigma} + R\left(w\frac{\theta}{\sigma}\right)}{\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}\right) \end{aligned}$$

(while can obtain better bounds, for our purposes, this proves adequate)

Moreover, the probability of forced stopping is

$$\begin{aligned} \Pr\left(\mathbf{m_i} < \mu + \sigma\theta R\left(w\frac{\theta}{\sigma}\right)\right)^N &= \Pr\left(\frac{\mathbf{m_i} - \mu}{\sigma\theta} < R\left(w\frac{\theta}{\sigma}\right)\right)^N \\ &= \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)^N \end{aligned}$$

Hence, the unconditional value Π upon stopping (forced or voluntary) is strictly in the interval given below.

$$\mu + \frac{\sigma}{\theta}R\left(w\theta/\sigma\right) + \frac{w}{1-\Phi\left(R\left(w\theta/\sigma\right)\right)} - \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)^N(K) < \Pi < \mu + \frac{\sigma}{\theta}R\left(w\theta/\sigma\right) + \frac{w}{1-\Phi\left(R\left(w\theta/\sigma\right)\right)}$$

where K is a constant independent of N. As may be noted, since $\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)^N \to 0$ when $N \to \infty$, we have that as N increases, Π converges (exponentially fast) to $\mu + \frac{\sigma}{\theta}R\left(w\theta/\sigma\right) + \frac{w}{1-\Phi(R(w\theta/\sigma))}$

Recall from proposition 3 that the probability of testing k (or more alternatives) is

$$\begin{aligned} P_k &= 1_{\{p_0 \leq R_k\}} \Pi_{j=1}^{k-1} \Pr\left(\mathbf{Z_j} \leq \frac{\sqrt{\sigma^2 + 1/\tau^2}}{\sigma^2}(R-\mu)\right) \\ &= 1_{\{p_0 \leq R_k\}} \Pr\left(\mathbf{Z_j} \leq R\left(w\frac{\theta}{\sigma}\right)\right)^{k-1} \end{aligned}$$

Hence, the expected number of evaluations before the firm stops is given by

$$P_1 + P_2 + \cdots = \frac{1}{1 - \Pr\left(\mathbf{Z_j} \leq R\left(w\frac{\theta}{\sigma}\right)\right)} = \frac{1}{1 - \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}$$

Hence, the expected total evaluation cost for the firm is given by:

$$\frac{1}{1 - \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)} \times w = \frac{w}{1 - \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}$$

Note: With finite N, we can modify the above calculation of total evaluation to obtain the expected number of evaluations as $\frac{1-\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)^N}{1-\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}$, and the expected costs as $w\frac{1-\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)^N}{1-\Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}$. As with the earlier case, we may note that these converge exponentially fast to the cost stated in the proposition. Lastly, note that as the cost and revenues both converge exponentially fast to the stated values, the profits will converge exponentially fast as well.

The last part of the proposition characterizing the optimal measurement technology is proved by noting that the net profits are given by

$$\mu + \left(\frac{\sigma}{\theta}R\left(w\theta/\sigma\right) + \frac{w}{1 - \Phi\left(R\left(w\theta/\sigma\right)\right)}\right) - \frac{w}{1 - \Phi\left(R\left(w\frac{\theta}{\sigma}\right)\right)}$$

and thus

$$\tau^* = \arg\max_{\tau}\left\{\mu + \frac{\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}}R\left(\frac{\sqrt{\sigma^2 + 1/\tau^2}}{\sigma^2}w\left(\tau\right)\right)\right\}$$

To show the sensitivity of τ^* with σ, it is sufficient to show that $\frac{\partial^2}{\partial\sigma\partial\tau}\left(\frac{\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}R\left(\frac{\sqrt{\sigma^2+1/\tau^2}}{\sigma^2}w\left(\tau\right)\right)\right) > 0$ for σ small enough, and < 0 for σ large.

For small σ, (under the assumption $w\left(\tau\right) = c\tau$), we have that

$$A\left(\sigma,\tau\right) = \frac{\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}}R\left(\frac{\sqrt{\sigma^2 + 1/\tau^2}}{\sigma^2}w\left(\tau\right)\right) \approx \sigma^2\tau R\left(\frac{c\tau^{\alpha-1}}{\sigma^2}\right) = B\left(\sigma,\tau\right)$$

It can be easily shown that $B(\sigma,\tau)$ is supermodular in (σ,τ). Similarly, for very large σ, we have that

$$A(\sigma,\tau) \approx \sigma R\left(w(\tau)\frac{1}{\sigma^1}\right) = C(\sigma,\tau)$$

We can also show that $C(\sigma,\tau)$ which is submodular in (σ,τ).

Hence, we have that τ^* is increasing in σ for small σ, and decreasing in σ for large σ. ■

Proof of Proposition 6

The firm's expected profit, after observing m_1, is given by

$$E\left[\max\left\{\mu_1+(m_1-\mu_1)\frac{\sigma_1^2}{\sigma_1^2+(1/\tau_1)^2},\mu_2+\frac{\sigma_2^2}{\sqrt{\sigma_2^2+(1/\tau_2)^2}}\mathbf{Z_2}\right\}\right]-w(\tau_2)$$

For notational convenience, define $\Delta_1=\mu_1+(m_1-\mu_1)\frac{\sigma_1^2}{\sigma_1^2+(1/\tau_1)^2}-\mu_2$, and define $\kappa_i=\frac{1}{\tau_i}$. Also, we shall maximize w.r.t κ and then obtain $\tau^*=\frac{1}{\kappa^*}$.

Thus, the firm's expected profit

$$E\left[\max\left\{\Delta_1,\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\mathbf{Z_2}\right\}\right]+\mu_2-w(1/\kappa_2)$$

[While not needed to prove the proposition, it is interesting to note that $\tau^*(\Delta_1)$ must be a symmetric function about the y-axis. Specifically, some algebra verifies that $E\left[\max\left\{-\Delta_1,\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\mathbf{Z_2}\right\}\right]=-\Delta_1+E\left[\max\left\{\Delta_1,\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\mathbf{Z_2}\right\}\right]$. Hence, the optimal κ_2^* (and $\tau_2^*=1/\kappa_2^*$) corresponding to Δ_1 must be identical to optimal τ_2^* (and $\tau_2^*=1/\kappa_2^*$) corresponding to $-\Delta_1$.]

Now consider, $E[\max\{x,\mathbf{Z}\}]$. If $R(y)=x$, then $E[\max\{x,\mathbf{Z}\}]-x=y$. That is, $E[\max\{x,\mathbf{Z}\}]=x+y=x+R^{-1}(x)$. Hence,

$$\begin{aligned}
\Pi(\Delta_1,\kappa_2) &= E\left[\max\left\{\Delta_1,\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\mathbf{Z_2}\right\}\right]+\mu_2-w(1/\kappa_2)\\
&= \frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}E\left[\max\left\{\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1,\mathbf{Z_2}\right\}\right]+\mu_2-w(1/\kappa_2)\\
&= \frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1+R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)+\mu_2-w(1/\kappa_2)\\
&= \Delta_1+\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)+\mu_2-w(1/\kappa_2)
\end{aligned}$$

Consequently,

$$\begin{aligned}\frac{\partial}{\partial\Delta_1}\left(\Pi\left(\Delta_1,\kappa_2\right)\right) &= 1+\frac{\sigma_2^2}{\sqrt{\sigma_2^2+\kappa_2^2}}\frac{1}{R'\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)}\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\\ &= 1+\frac{1}{R'\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial\kappa_2\partial\Delta_1}\left(\Pi\left(\Delta_1,\kappa_2\right)\right) &= \frac{-1}{\left(R'\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)\right)^2}R''\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)\times\\ &\quad\frac{1}{R'\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)}.\frac{1}{2}.\frac{2\kappa_2}{\sigma_2^2\sqrt{\sigma_2^2+\kappa_2^2}}\Delta_1\\ &= \frac{-1}{\left(R'\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right)\right)^3}R''\left(R^{-1}\left(\frac{\sqrt{\sigma_2^2+\kappa_2^2}}{\sigma_2^2}\Delta_1\right)\right).\frac{\kappa_2}{\sigma_2^2\sqrt{\sigma_2^2+\kappa_2^2}}\Delta_1\end{aligned}$$

Since $R''>0$ and $R'<0$, we have that $\frac{\partial^2}{\partial\kappa_2\partial\Delta_1}\Pi\left(\Delta_1,\kappa_2\right)>0$ if $\Delta_1>0$; and $\frac{\partial^2}{\partial\kappa_2\partial\Delta_1}\Pi\left(\Delta_1,\kappa_2\right)<0$ if $\Delta_1<0$. Thus, κ^* is increasing (and thus τ^* is decreasing) in Δ_1 when $\Delta_1>0$, and κ^* is decreasing (and thus τ^* is increasing) in Δ_1 otherwise. ■

Proof of Proposition 5′

With $N=1$, if the firm employs a measurement technology τ at cost $w\left(\tau\right)$, then the expected payoffs are given by $E\left[\max\left\{0,\mu+\mathbf{Z}\frac{\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}\right\}\right]-w\left(\tau\right)$. Some algebra verifies that the revenues may be simplified as

$$E\left[\max\left\{0,\mu+\mathbf{Z}\frac{\sigma^2}{\sqrt{\sigma^2+1/\tau^2}}\right\}\right]=\left(\sigma/\theta\right)\left(\frac{\mu\theta}{\sigma}\Phi\left(\frac{\mu\theta}{\sigma}\right)+\frac{e^{-\frac{\mu^2\theta^2}{2\sigma^2}}}{\sqrt{2\pi}}\right)$$

where $\theta=\sqrt{1+1/(\sigma\tau)^2}$.

When $\mu=0$, we have that the payoffs is given by $\left(\sigma/\theta\right)\left(\frac{1}{\sqrt{2\pi}}\right)-w\left(\tau\right)$. Hence, optimal $\tau^*=\arg\max_\tau\left\{\left(\sigma/\theta\right)\left(\frac{1}{\sqrt{2\pi}}\right)-w\left(\tau\right)\right\}$. Note that $\frac{\partial^2}{\partial\sigma\partial\tau}\left(\frac{\sigma}{\theta}\right)>0$ for small σ, and is negative otherwise. Hence, τ^* is increasing in σ for low enough σ, and is decreasing in σ for high enough σ. ■

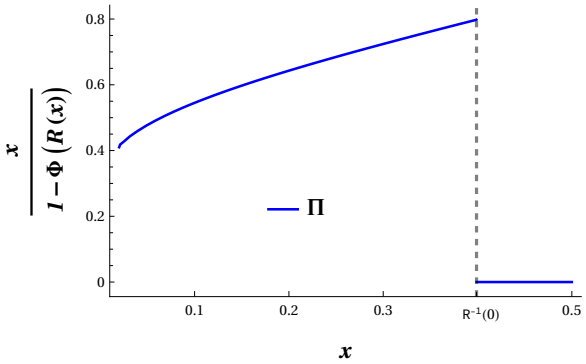


Figure 8: $x\frac{1}{1-\Phi(R(x))}$ as a function of x for $x < R^{-1}(0)$

Proof of Proposition 7

With the platform offering measurement technology τ at price w, we have that their profits (when the client firm follows the optimal policy as outlined previously) is given by

$$\Pi(\tau, w(\tau)) = \begin{cases} \dfrac{w-c}{1-\Phi\left(R\left(w\theta/\sigma\right)\right)} & \text{if } w\theta/\sigma \leq R^{-1}(0) \\ 0 & \text{o/w} \end{cases}$$

Define $x = w\theta/\sigma$. Then, the profits can be rewritten as

$$\Pi = \begin{cases} \left(\dfrac{\sigma}{\theta}\right)\left(x - c\left(\dfrac{\theta}{\sigma}\right)\right)\dfrac{1}{1-\Phi\left(R\left(x\right)\right)} & \text{if } x < R^{-1}(0) \\ 0 & \text{o/w} \end{cases}$$

We focus our attention in the region $x < R^{-1}(0)$ since that is where the platform's profits are greater than 0. In this region, if $\frac{1}{1-\Phi(R(x))} + x\frac{d}{dx}\left(\frac{1}{1-\Phi(R(x))}\right) \geq 0$, then $\frac{1}{1-\Phi(R(x))} + (x - c\theta/\sigma)\frac{d}{dx}\left(\frac{1}{1-\Phi(R(x))}\right) > 0$ for $c\theta/\sigma \leq x \leq R^{-1}(0)$, and the optimal $x^* = R^{-1}(0)$. Thus, to prove that Π is maximized at $x = R^{-1}(0)$, it is enough to prove that $\frac{1}{1-\Phi(R(x))} + x\frac{d}{dx}\left(\frac{1}{1-\Phi(R(x))}\right) \geq 0$, i.e., that $x\frac{1}{1-\Phi(R(x))}$ is increasing in x for $x < R^{-1}(0)$. Figure 8 plots $x\frac{1}{1-\Phi(R(x))}$, and shows that it is increasing in x, and is indeed maximized at $x = R^{-1}(0) \approx 0.398$. Thus, the optimal

$$\begin{aligned} w &= \left[\frac{\sigma}{\theta}\right] R^{-1}(0) \\ &= \frac{R^{-1}(0)\,\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} \end{aligned}$$

Proof of Proposition 8

From proposition 7, we can substitute the optimal fee to obtain the platform's profit. These profits can be represented as as

$$\Pi(\tau) = \frac{R^{-1}(0)\,\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} - c(\tau)$$

So, the optimal fidelity chosen by the platform is given by

$$\tau_p^* = \arg\max_{\tau} \left\{ \frac{R^{-1}(0)\,\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} - c(\tau) \right\}$$

Note that

$$\frac{\partial^2}{\partial\tau\partial\sigma}\left(\frac{R^{-1}(0)\,\sigma^2}{\sqrt{\sigma^2 + 1/\tau^2}} - c(\tau)\right) = \frac{R^{-1}(0)\,\sigma\left(2 - \sigma^2\tau^2\right)}{\tau\sqrt{\sigma^2 + \frac{1}{\tau^2}}\left(\sigma^2\tau^2 + 1\right)^2}$$

which is positive for sufficiently low σ, and negative for sufficiently high σ. Hence, τ^* is increasing in σ for sufficiently low σ, and decreasing in σ for sufficiently high σ ■