# Optimizing Service Failure and Damage Control*

Daniel Halbheer, Dennis Gärtner, Eitan Gerstner, and Oded Koenigsberg†

November 2017

**Abstract**

Should a provider deliver a reliable service or should it allow for occasional service failures? This paper derives conditions under which randomizing service quality can benefit the provider and society. In addition to cost considerations, heterogeneity in customer damages from service failures allows the provider to generate profit from selling damage prevention services or offering compensation to high-damage customers. This strategy is viable even when reputation counts and markets are competitive.

*Keywords:* Service Quality, Service Reliability, Service Failure, Damage Control.

*We thank Roland Rust and seminar participants at the EMAC 2016 (Oslo), the INFORMS Marketing Science Conference 2016 (Shanghai), and the Thought Leaders in Service Marketing Strategy Conference 2016 (Paris) for helpful comments and suggestions. In addition, we thank the review team for their helpful comments, which greatly improved the paper.

†Daniel Halbheer: HEC Paris, Marketing Department, 1 rue de la Libération, 78351 Jouy-en-Josas Cedex, France (halbheer@hec.fr). Dennis Gärtner: University of St. Gallen, Institute of Economics (FGN), Varnbüelstrasse 19, 9000 St. Gallen, Switzerland (dennis.gaertner@unisg.ch). Eitan Gerstner: Technion – Israel Institute of Technology, Faculty of Industrial Engineering and Management, Technion City, Haifa 32000, Israel (gerstner@technion.ac.il). Oded Koenigsberg: London Business School, Regent's Park, London, NW1 4SA, United Kingdom (okoenigsberg@london.edu). Daniel Halbheer is a member of the CNRS unit GREGHEC, UMR CNRS 2959, and gratefully acknowledges the support of the HEC Foundation. Corresponding author: Daniel Halbheer.

# 1 Introduction

The meeting has just ended on time and you are ready to head out to the airport to catch your flight, so you open the *Uber* app. After typing the destination into the 'Where to' box, the app requires you to make a choice among two economy options: *uberX* at a price of €50.00 or *POOL* at a price of €35.00, as illustrated in Figure 1. The option *uberX* offers private rides, while the option *uberPOOL* matches riders headed into the same direction.[1] Evidently, the sharing option saves costs but may increase the ride time, which creates inconvenience and a damage if you arrive late. Which option would you select? Clearly, the choice among the consistent service *uberX* and the potentially less reliable service *uberPool* depends on the likelihood that other riders will join the ride, and the damage in case you miss your flight.

The novel aspect of this service differentiation strategy is that the provider offers a randomized service quality alongside a consistent service (Parasuraman, Zeithaml, and Berry 1988, 1985).[2] Randomization as part of the business strategy is common in pricing. Examples include probabilistic or opaque selling (Zhang, Joseph, and Subramaniam 2015; Fay and Xie 2008), participative pricing (Chen, Koenigsberg, and Zhang 2016; Schmidt, Spann, and Zeithammer 2014), and randomized pricing (Varian 1980). This naturally raises the question: Can service providers and society likewise benefit from offering a randomized service quality?

To find answers, we present a model in which a provider designs the service by choosing the failure rate and the price. Service failures are assumed to have a dual impact: they lead to cost savings for the provider, but they inflict damages on customers. Customer damages from service failures may be monetary (direct follow-up costs) or non-monetary (such as the opportunity cost of time, or hassle costs), or both.[3] Clearly, reducing the failure rate would lead to lower expected damages for customers. However, this increases the cost of providing the service. This tradeoff between size and frequency of customer

[1] For details, see www.uber.com.

[2] *United* recently launched its new *Flex-Schedule Program*, which allows opted-in passengers to benefit from vouchers up to $250 if they are willing to accept changes of the flight time within a given day or a downgrade from Economy Plus to regular-old Economy (Bloomberg 2017).

[3] By choosing a strictly positive failure rate, the provider randomizes service quality and imposes expected damages on its customers—a "calculated misery" (Wu 2014).

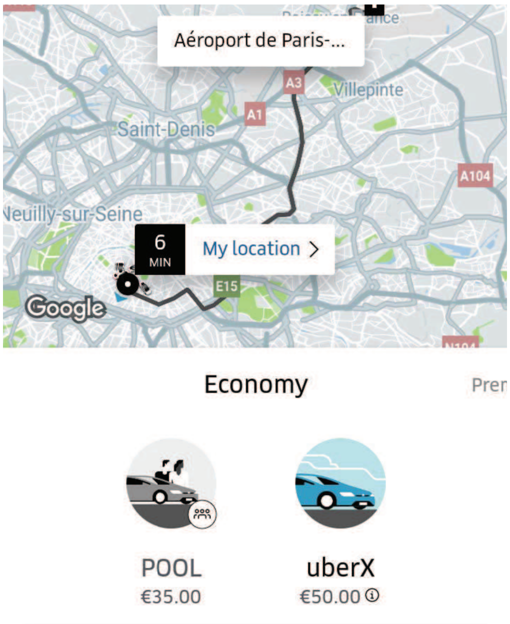


**Figure 1:** Screenshot of the *Uber* app.

damages and the cost of reducing service failures drives the optimal failure rate, and therewith the optimal service quality. The model is used to determine the privately and socially optimal failure rates and prices under different market conditions, captured by (a) the cost structure (technology) of the provider, (b) the size of customer damages, and (c) by the degree of customer heterogeneity regarding damage tolerance. In addition, the model is used to examine optimal strategies to enhance profit by offering failure prevention (service backups and protection plans) and monetary damage compensation.[4] Figure 2 summarizes the main elements of the model.

Several key results are provided. First, randomizing service quality can be optimal for the provider and society more broadly. Reducing the failure rate to zero is optimal and economically efficient only if the benefit of eliminating the customer damages exceeds the cost of fail-safing the service. Second, customer heterogeneity in damage tolerance reinforces the incentives to retain service failures but results in an economically inefficient

[4]The idea of failure prevention is to address service failures before customers become aware of any damage. For example, providers can apply fail-safing methods (poka-yokes) such as fishbone diagrams or Pareto charts to identify and eliminate failure points (Chase and Stewart 1994). In contrast, compensation becomes relevant once the damage has occurred, and therefore tends to be inefficient, especially when the damage exceeds the price paid for the service.

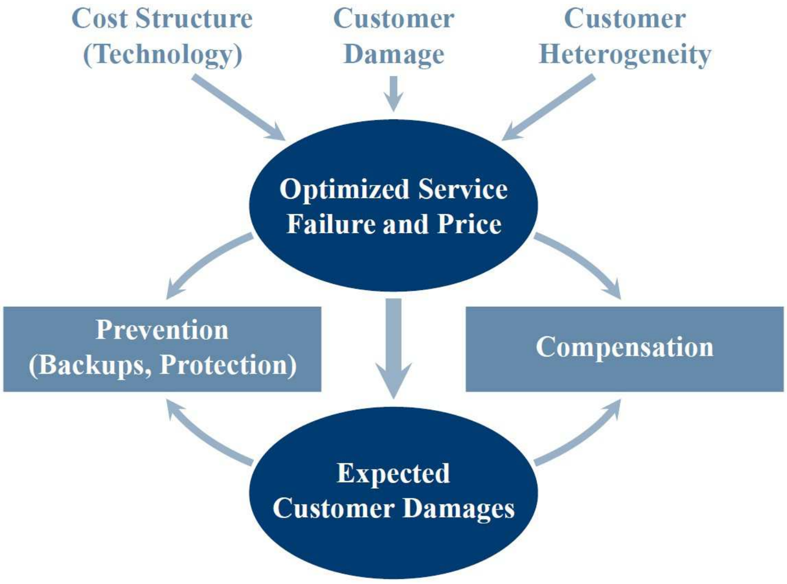


**Figure 2:** Model Elements.

outcome. Third, use of failure-prevention strategies such as backups and protection strengthen the provider's incentives to offer an unreliable service, whereas the optimal failure rate is unaffected by offering an ex-post compensation strategy. Fourth, we show that offering protection services as a form of selective service guarantee allows the provider to tap into a new source of revenue from customers who wish to avoid damages. Finally, competitive forces neither drive a provider of unreliable service out of the market, nor do they lower the optimal failure rate. Hence, counter to intuition, rivalry in the market need not necessarily improve service quality.

These results contribute to the service literature in several ways. First, we introduce the notion of "randomized service quality" into the service quality literature (Chase and Stewart 1994; Parasuraman, Zeithaml, and Berry 1988, 1985; Levitt 1972). The quality of the service is randomized in the sense that it is characterized by a Bernoulli distribution governed by the failure rate rather than by a deterministic quality level. Our analysis builds on the "return on quality" approach (Rust, Zahorik, and Keiningham 1995; Rust and Zahorik 1993) and shows that eliminating service failures is not necessarily optimal. This also fits in line with recent work that shows that there is an optimal level of service productivity (Rust and Huang 2012) or that it may not be profitable to keep or redress all customers (Dukes and Zhu 2016; Ma, Sun, and Kekre 2015; Shin and Sudhir 2015). Importantly, we extend this line of research by providing a welfare analysis that helps managers to understand the societal impact of purposely introducing service failures that lead to customer damages. To put the notion of randomized service quality into perspective, it is crucial to highlight the difference to probabilistic selling, a strategy under which the seller creates probabilistic goods using the seller's existing distinct products or services (Fay and Xie 2008).[5] In contrast, we study design decisions of services by letting the provider choose the failure rate and the price, but do not allow for probabilistic selling. Thus, whenever the provider optimally decides to retain service failures, the quality of the service is randomized.

Second, we introduce the notion of "random versioning" into the product line literature (Deneckere and McAfee 1996; Shapiro and Varian 1996; Moorthy and Png 1992) and show under what conditions such a sales strategy is socially harmful. The novelty is that the provider offers services with randomized quality rather than selling different deterministic qualities at different prices. In fact, there are instances where some customers may end up getting a reliable service as those who actually pay for it, but pay a lower price because of the randomized service quality. Such a probabilistic upgrade to a higher service tier is not a possibility in traditional models of quality differentiation. Under probabilistic selling in quality-differentiated markets (Zhang, Joseph, and Subramaniam 2015), such a quality upgrade is possible, albeit for a different reason: By offering a synthetic service consisting of a lottery between two distinct services, some customers with a low valuation will be randomly assigned to obtain the high-quality service. We will show below that the notion of random versioning is useful to provide an explanation for the probabilistic nature of the service *uberPool*.

Third, by simultaneously studying failure optimization and damage control, we add to the literature on service guarantees (Hogreve and Gremler 2009). A service guarantee is a promise by the provider to deliver a certain level of service and offer a compensation in the event of a service failure.[6] Our key insight is that the decisions about the level of service

[5]Initially introduced in the context of horizontal markets, the profit impact of probabilistic selling has recently been studied in the context of quality-differentiated markets (Zhang, Joseph, and Subramaniam 2015).

[6]Chen, Gerstner, and Yang (2012, 2009) study monetary compensation as a form of service recovery under exogenous service failures.

(the failure rate) and the design of the damage control (prevention and compensation) are intertwined, and thus should be made together by adopting a global view in the service planning process. Recognizing this fact helps providers to understand that the instruments of damage control in effect become drivers of service failure—similar to costs, damages, and customer heterogeneity.[7] This third contribution is unique to the service literature because our analysis acknowledges long-recognized differences between the nature of service and the nature of products (Rust and Chung 2006; Parasuraman, Zeithaml and Berry 1985). Specifically, the characteristic of "perishability"—once the time of service passes, the opportunity to deliver the service at that time is no longer available—drives the key differences to the literature in the product realm: Service backups (nondiscriminatory failure prevention) and damage protection (selective failure prevention) are simply not an option. Our paper also differs from the literature on planned obsolescence, which studies ways to limit the physical life of durable products (Waldman 1993; Bulow 1986). We examine the drivers of service failures and suggest economically efficient ways to manage them.

The remainder of the paper is organized as follows. Section 2 analyzes the role of the cost structure (technology) and the customer damages for the optimal choice of the failure rate and the price. Section 3 extends the basic model and studies how customer heterogeneity in damage tolerance affects the optimal design of the service. Section 4 shows when it is optimal for the provider to use a backup strategy. Section 5 shows that marketing protection plans helps to tap into a new source of revenue. Section 6 shows how to design the optimal compensation strategy. Section 7 studies the impact of competition. Section 8 concludes and outlines managerial implications.

# 2 Optimized Service Failure

This section derives the optimal failure rate of the service and shows how it is driven by the interplay of the cost structure (technology) of the provider and the customer damages. We consider a monopoly provider who offers a probabilistic service that can either fail or

[7] More broadly, this result can also be used to explain the observed dispersion of failure rates within and across service industries, and provide insights into the questions of why (intentionally) poor service prevails in the marketplace (Gerstner and Libai 2006).

**Table 1:** Service Outcomes.

| Service delivery at price $p$ | Outcome | Probability | Customer damage | Realized utility |
|---|---|---|---|---|
| Meets expectations | Success | $1-q$ | 0 | $v-p$ |
| Fails to meet expectations | Service failure | $q$ | $d>0$ | $v-p-d$ |

succeed to potential customers. A service failure occurs when the service does not meet customer expectations. Conversely, the service succeeds when the expectations are met (Parasuraman, Zeithaml, and Berry 1985). Table 1 gives an overview of the model with probabilistic service outcomes.

The provider designs the service by choosing the failure rate $q$ and the price $p$.[8] The choice of the failure rate $q$ directly determines the reliability rate $1-q$: If $q=0$, the provider offers a (perfectly) reliable service and prevents service failures. Instead, when setting $q>0$, the provider offers an unreliable service and retains service failures.

The unit cost of providing the service at any given failure rate $q$ is $c(q)$. We assume that $c'(q)<0$ and $c''(q)>0$, which means that reducing the failure rate is costly, and more so the lower the failure rate is. Offering a reliable service costs $c(0)\leq\infty$, which allows for the possibility that eliminating service failures (i.e. fail-safing) is prohibitively costly. The constant unit-cost formulation assumes that the total cost to the provider is linear in the number of customers being served (no economies of scale).

There is a unit mass of potential customers in the market, who are identical and risk neutral.[9] Each has valuation $v$ for the service and faces a damage $d$ in the event of a service failure. The preference parameters $v$ and $d$ are strictly positive and known to the provider. Potential customers know the failure rate $q$ and the price $p$ when deciding whether to purchase the service or choose an outside option, the utility of which is normalized to zero. The expected utility of the service is

$$u(q,p)=v-qd-p.$$

[8]Ownership and control are assumed to coincide, which lets us abstract from agency issues such as formulating quality goals for management.

[9]We relax both assumptions in turn below, when considering heterogeneous customers in Section 3, and when considering risk-averse customers in Section 6.

Thus a higher failure rate (a lower reliability rate) and a higher price reduce the expected utility of the service. It will be useful to define $v - qd$ as the expected service quality—a definition that is consistent with Parasuraman, Zeithaml, and Berry (1985), who show that reliability is a key determinant of (expected) service quality. The service quality itself follows a Bernoulli distribution with an ex ante known failure rate (a search attribute) and an ex post realization (an experience attribute).[10] A potential customer then purchases the service if the price does not exceed the expected quality—even though it is possible that the realized damage exceeds the valuation of the service.[11]

**Optimal Decisions.** The provider chooses the failure rate $q$ and the price $p$ of the service in order to maximize the expected profit subject to the constraints that the potential customers participate in the market and that $q$ is a well-defined probability:

$$\begin{aligned} \max_{q,p} \quad & \pi(q,p) = p - c(q) \\ \text{s.t.} \quad & p \leq v - qd \\ & 0 \leq q \leq 1. \end{aligned} \tag{1}$$

It is optimal to set the price equal to the expected quality, that is, $p^* = v - qd$. This allows the provider to fully extract the surplus from the customers. The following result holds:

**Proposition 1 (Optimized Service Failure).** *It is optimal and economically efficient to retain service failures if $-c'(0) > d$ and to offer a reliable service if $-c'(0) \leq d$.*

Proposition 1 mirrors the "return on quality" approach (Rust, Zahorik, and Keiningham 1995): If the marginal cost to fail-safe the service $-c'(0)$ is lower than or equal to the customer damage $d$, then there is a corner solution $q^* = 0$ (reliable service). Intuitively, the increase in pricing power that results from the elimination of customer damages outweighs the higher cost of the service. Instead, if the marginal cost to fail-safe the service exceeds the damage, then the optimum has $q^* \in (0,1)$ (unreliable service). Note that the choices of the provider maximize welfare $W(q) \equiv v - qd - c(q)$—the sum of profit and consumer

[10]The assumption that $q$ is known to potential customers implies that the provider does not face a reputation issue: The failure rate is a search attribute that is known prior to purchasing, whereas the delivery itself is an experience attribute that can only be learned after purchase (Nelson 1974).

[11]For simplicity, we model service provision as a static (one-shot) decision, but costs and damages may of course nevertheless be thought of as compounding all future costs and damages.

surplus—because the surplus from the customers can be fully extracted by pricing to expected value. The economic approach therefore shows that the "return on quality" approach is optimal not only from the perspective of the provider but also from a societal point of view.[12]

An important managerial implication of Proposition 1 is that it can be optimal for the provider to randomize service quality rather than opting for "consistency of performance" (Parasuraman, Zeithaml, and Berry 1988, 1985). Specifically, if $q^* \in (0,1)$, service quality is characterized by a (Bernoulli) distribution rather than by a deterministic quality level.[13] The following result shows how the optimal failure rate $q^*$ is affected by changes in the business environment.

**Lemma 1 (Adapting to Change).** *Higher marginal cost of fail-safing or lower customer damages increase the optimal failure rate if* $-c'(0) > d > -c'(1)$.

Intuitively, a higher marginal cost to fail-safe the service motivates the provider to increase the failure rate and thereby save costs—even though this reduces the expected quality and thus the pricing power. Figure 3 illustrates this comparative statics result and highlights the "fail-safing zone" where $q^* = 0$. Similarly, a lower damage $d$ increases the expected quality and the pricing power. To counterbalance the higher pricing power, it is optimal to increase the failure rate $q^*$.

Lemma 1 shows how a provider should adjust the failure rate in response to a change in cost. To make this (implicit) dynamic process more transparent, we consider a scenario where an existing technology $c_0(q)$ can be replaced with a new technology $c_1(q)$ at a fixed cost $F > 0$. Adopting the new technology is optimal if

$$d(q_0^* - q_1^*) - [c_1(q_1^*) - c_0(q_0^*)] > F, \tag{2}$$

that is, if the change in revenue net of the change in variable costs exceeds the adoption cost $F$. This shows that financial benefits from restructuring may be derived from revenue

[12]This analysis assumes that society seeks to maximize total expected surplus, which seems plausible in many but perhaps not all situations. Not least, randomization will create different (ex post) outcomes for otherwise identical individuals, which may run counter to certain fairness goals, particularly when damages are "large."

[13]If $q^* = 1$, the service fails deterministically and is perfectly unreliable. In essence, this is a quality-degraded version of the reliable service.

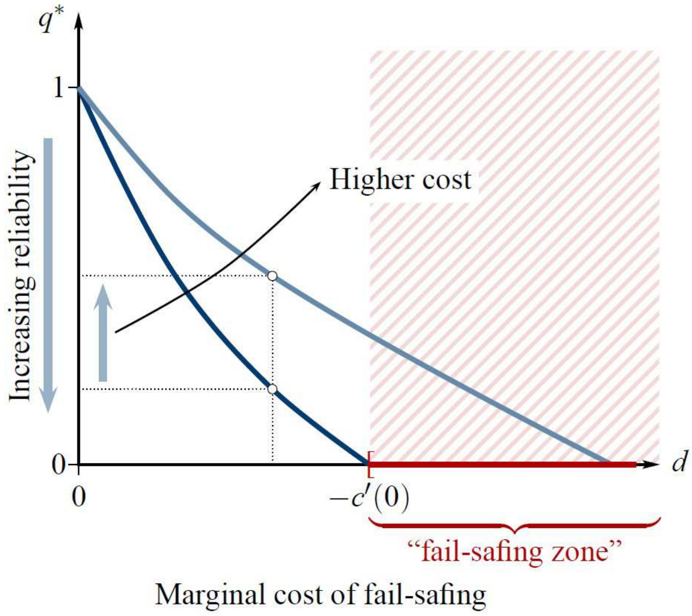


**Figure 3:** Optimal failure rate $q^*$ as a function of the damage $d$.

expansion (if $q_1^* < q_0^*$), cost reduction (if $c_1(q_1^*) < c_0(q_0^*)$), or both simultaneously (Rust, Moorman, and Dickson 2002).

**Example.** To illustrate our results, we consider the cost function $c(q) = c_0(1-q)^2$, where $c_0 > 0$ is the cost to fail-safe the service. If $d < 2c_0$, it is optimal for the provider to deliver an unreliable service (Proposition 1). The optimal failure rate is given by

$$q^* = 1 - \frac{d}{2c_0}.$$

This shows that a higher marginal cost to eliminate service failures, given by $2c_0$, leads to a higher optimal failure rate (Lemma 1). A higher damage $d$ has the opposite effect: It increases the marginal benefit of reducing $q$ and encourages the provider to incur higher costs by choosing a lower $q^*$.

To illustrate the adoption of a new technology at a fixed cost $F > 0$, suppose that $c_0(q) = c_0(1-q)^2$ and that $c_1(q) = c_1(1-q)^2$, where $c_1 < c_0$ reflects the technological progress (efficiency gains through lower costs). Based on (2), adopting the new technology has financial benefits for the provider if

$$F < \frac{(c_0 - c_1)d^2}{2c_0c_1}.$$

Note that higher efficiency gains and higher customer damages make the adoption of the new technology more attractive for the provider.

# 3 Customer Heterogeneity

This section shows that customer heterogeneity can be another driver of service failure and give rise to socially harmful "random versioning." To capture customer heterogeneity, we assume that there are two types of potential customers, 'lows' ($i = L$) and 'highs' ($i = H$), who differ in their valuations $v_i$ for the service and the damages $d_i$ that result from a service failure. The expected utility of the service for type $i$ is

$$u_i(q, p) = v_i - q\,d_i - p,$$

whereas the utility of the outside option is normalized to zero. The parameters $v_i$ and $d_i$ are strictly positive, and we assume that $\Delta v \equiv v_H - v_L > 0$ and $\Delta d \equiv d_H - d_L > 0$, so that highs have not only higher valuation for a successful service, but also suffer higher damage from a failure. In addition, we assume that $\Delta v \geq \Delta d$, which ensures that the latter never dominates, in the sense that any given service with arbitrary failure rate will still be valued more by highs than by lows. Finally, we let $\alpha \in (0, 1)$ denote the proportion of highs in the population and assume that the provider cannot observe individual types but only knows the proportion $\alpha$.

**Optimal Decisions.** The provider offers a segment-specific failure rate $q_i$ and price $p_i$, chosen so as to maximize expected profit subject to each type preferring 'their' offer over the outside option (the participation constraints PC$_i$) and the other type's offer (the self selection or incentive constraints IC$_i$):[14]

$$\max_{\{q_i, p_i\}_{i=L,H}} \quad \pi(\mathbf{q}, \mathbf{p}) = (1-\alpha)(p_L - c(q_L)) + \alpha(p_H - c(q_H)) \tag{3}$$

$$\text{s.t.} \quad p_i \leq v_i - q_i\,d_i \tag{PC$_i$}$$

$$p_i \leq p_j + (q_j - q_i)d_i, \tag{IC$_i$}$$

[14]We focus on the interesting case where it is optimal for the provider to offer a differentiated service and sell to all potential customers. This involves no loss of generality so long as there are sufficiently many lows in the population. Details are given in the Appendix.

where $\mathbf{q} \equiv (q_L, q_H)$ and $\mathbf{p} \equiv (p_L, p_H)$. Given the provider's objective of extracting as much surplus from customers as possible to boost profit, the relevant constraints are the lows' participation constraint ($\mathrm{PC}_L$) and the highs' self-selection constraint ($\mathrm{IC}_H$), so that these constraints will bind in the optimum. The following result holds:

**Proposition 2 (Random Versioning).** *Offering a reliable service to customers with high damages and randomizing service quality for customers with low damages is optimal but economically inefficient if* $-c'(1) < d_L - \frac{\alpha}{1-\alpha}\Delta d < -c'(0) \leq d_L$.

This result is reminiscent of service differentiation in product line design (Moorthy and Png 1992), versioning (Shapiro and Varian 1996), and damaged goods (Deneckere and McAfee 1996). The novelty is that the provider may engage in "random versioning": Rather than selling two deterministic qualities at different prices, the service targeted at lows is probabilistic. Therefore, lows may end up getting a successful service as do highs, but pay a lower price because of the randomized service quality. Such a probabilistic upgrade to a higher service tier is not a possibility in traditional models of quality differentiation. Under probabilistic selling in quality-differentiated markets, customers who purchase the synthetic service may obtain the high-quality offering with some preannounced probability (Zhang, Joseph, and Subramaniam 2015). The difference to our model is that the upgrade results as an implication of the pricing strategy and not because of the service-design decision.

Engaging in randomized versioning is socially harmful. Proposition 2 shows that even when the two drivers customer damages and cost structure themselves would actually call for elimination of service failures (in the sense that elimination would be economically efficient), the presence of customer heterogeneity might lead the provider to instead target only highs with a reliable service, and lows with an unreliable service. The intuition for the economic inefficiency is that, when customers' valuations are privately known, the provider can no longer extract the full surplus from customers. Rather, to keep highs from taking the service targeted at lows, the provider must surrender a so-called 'information rent' to highs. This rent can be reduced by increasing $q_L$ above the efficient failure rate $q^*$, as this makes the lows' offer less attractive to highs. The flip side of doing so is that this not only creates a distortion of the failure rate offered to lows but also excessive expected damages. If the provider were to engage in socially responsible business practices instead, randomizing service quality for the lows might in fact be considered unethical behavior.

Proposition 2 helps to shed light on the *Uber* example. Consistent with our prediction, the service *uberX* targeted at highs (at a high price) is not randomized: a rider is offered a solo ride for sure (deterministic service quality). In contrast, the service targeted at lows (at a low price) is randomized: a rider typically shares the ride but may be offered a solo ride (randomized service quality). From the perspective of a customer who selects the option *Pool*, the solo ride is a "success." The example shows that optimizing the failure rate through the design of the matching algorithm is definitely a core element of the business strategy for a company like *Uber*.[15]

The next result shows how customer heterogeneity drives the adverse societal impact of random versioning.

**Lemma 2** (**Calculated Misery**). *A larger proportion of highs or a larger difference in damages increase the economic inefficiency and therefore the expected damages for lows.*

Intuitively, an increase in customer heterogeneity motivates the provider to increase the failure rate for lows to reduce the information rent $\alpha(\Delta v - \tilde{q}_L \Delta d)$ that accrues to highs. Put differently, the provider has an incentive to increase $\tilde{q}_L$ in response to a higher $\alpha$ or $\Delta d$ to leave less money on the table for highs. The excessively high failure rate to boost profit therefore creates a truly calculated misery for lows.

**Example.** To illustrate our results, consider the cost function $c(q) = c_0(1-q)^2$ and suppose that $0 < d_L - \frac{\alpha}{1-\alpha}\Delta d < 2c_0 \leq d_L$. The optimal failure rates are

$$\tilde{q}_L = 1 - \frac{d_L - \frac{\alpha}{1-\alpha}\Delta d}{2c_0} \quad \text{and} \quad q_H^* = 0,$$

whereas it would be economically efficient to offer a reliable service to all customers, that is $q_L^* = q_H^* = 0$. The extent of the upward distortion is positively related to the proportion of highs $\alpha$ and the difference in damages $\Delta d$.

[15]The lack of data on the cost structure and customer valuations does not allow a judgement on whether or not the marketing strategy is socially harmful.

# 4 Service Backups

Service providers often use contingent planning and rely on backups to prevent customer damages that result from service failures (Gallego and Stefanescu 2012; Biyalogorsky and Gerstner 2004). This section shows when it is optimal to offer service backups .

We assume that the provider can purchase backups at a unit price $w$ on a wholesale market. In the event of a service failure, a backup can be used by the provider to nonetheless deliver a reliable service and prevent customer damages. Effectively, this means that the provider has some lead time to address service failures before the customers become aware of any damage—a form of a "service guarantee." The provider can choose among two strategies: offer the service with or without backups, where the latter corresponds to the benchmark case in Section 2.

**Optimal Decisions.** When using backups, the provider chooses the failure rate $q_b$ and the price $p$ of the service in order to maximize the expected profit subject to the participation constraint of the potential customers:

$$\begin{aligned} \max_{q_b, p} \quad & \pi(q_b, p) = p - c(q_b) - q_b w \\ \text{s.t.} \quad & p \leq v. \end{aligned}$$

The provider sets the optimal price equal to the valuation of the service, that is, $p^* = v$. Consequently, the first-order condition for a profit-maximizing choice of the failure rate is $-c'(q_b^*) \leq w$, which again mirrors standard cost-benefit analysis: A marginal decrease in the failure rate $q$ increases the unit cost of the provider by $-c'(q)$ and reduces the expected cost of backups by $w$. Clearly, the optimal failure rate $q_b^*$ depends on the relative strength of these two drivers of service failure. The following result holds:

**Proposition 3 (Backups).** *Using service backups is optimal and economically efficient if $w < d$, but increases the failure rate $q_b^*$ above the optimal failure rate $q^*$ characterized in our benchmark case (Section 2). If $w \geq d$, the provider does not offer backups and the optimal failure rate is $q^*$.*

Proposition 3 shows that it is optimal for the provider to use backups if the wholesale price $w$ is lower than the damage $d$. If $w < d$, backups are economically efficient because their expected costs are lower than the expected damages that would result otherwise. Surprisingly, backups increase the optimal failure rate compared to the benchmark case. To grasp the intuition for this result, note that the marginal benefit of reducing $q$ is $w$ (lower procurement cost) with backups, while it is $d$ (higher pricing power) in the benchmark case. The lower marginal benefit of reducing $q$ leads the provider to increase the failure rate above the level that would be optimal absent service backups. Bearing this upward distortion of the failure rate in mind is important for managers when projecting the costs of a strategy with service backups.

**Example.** To illustrate the distortion of the failure rate in Proposition 3, consider the cost function $c(q) = c_0(1-q)^2$ and suppose that $-c'(0) > d$. The upward distortion of the failure rate is

$$q_b^* - q^* = \frac{d-w}{2c_0}.$$

This shows that the gap in the failure rates is proportional to $d-w$ and decreases in the marginal cost to fail-safe the service, which is given by $2c_0$. The latter simply reflects the fact that both $q_b^* \to 1$ and $q^* \to 1$ as $c_0 \to \infty$.

# 5 Protection Services

Service backups are nondiscriminatory in that they prevent damages for all customers. In this section, we study protection services that selectively prevent damages for customers who have purchased a protection plan.

We assume that the provider offers a basic service at price $p$ that fails with probability $q$. In addition to the basic service, the provider markets a protection plan as an add-on to the basic service at price $f$—the protection fee—that prevents service failures.[16] As in the case of backups, we suppose that the provider can address service failures before protected customers are aware of them—a form of a selective "service guarantee." The key difference to backups is (besides being selective) that protection services are supplied internally, whereas backups are procured externally on a wholesale market. The provider faces two types of potential customers—lows and highs—as described in Section 3.

[16]Examples include protection plans against contingent charges such as unauthorized overdraft fees for bank accounts or late payment fees for credit cards.

**Optimal Decisions.** The provider chooses the failure rate $q$ and the price $p$ for the basic service and the protection fee $f$ to maximize expected profit subject to the constraints that lows purchase the basic service and that highs purchase both the basic service and the protection:

$$\max_{q,p,f} \quad \pi(q,p,f) = (p - c(q)) + \alpha(f - [c(0) - c(q)]) \tag{4}$$
$$\text{s.t.} \quad p \leq v_L - q\,d_L$$
$$f \leq q\,d_H.$$

There are two sources of revenue: from selling the standard service to all customers and from selling protection to the highs. Therefore, offering protection allows the provider to tap into a new source of revenue from customers who wish to avoid damages.

**Proposition 4 (Protection Plans).** *Marketing protection services is optimal but socially harmful when the marginal cost to fail-safe the basic service is arbitrarily small.*

Proposition 4 shows that the provider has an incentive to segment the market and offer protection services even when the costs of providing a reliable basic service are negligible. The problem of this business practice—even though it maximizes the profit of the provider—is that it is detrimental to society: Offering a service with $q = 0$ at price $p = v_L$ instead would increase welfare.[17] This result shows that the incentive to offer an unreliable service is not driven by the cost structure alone: it is the temptation to exploit customers by extracting surplus to boost profit. Managers with a broader focus than pure profit orientation would solve the problem at the source rather than treating symptoms through protection plans.

**Example.** To illustrate, consider the cost function $c(q) = \varepsilon(1-q)^2$, where $\varepsilon > 0$ is a small but positive number. This specification entails that the marginal cost to fail-safe the service is essentially zero, that is, $c'(0) = \varepsilon$. If $\varepsilon \leq d_L$, the optimal failure rate for the basic service is given by

$$q = 1 - \frac{d_L - \frac{\alpha}{1-\alpha}\Delta d}{2\varepsilon}.$$

[17]In the limiting case where the marginal cost to fail-safe the service approaches zero, the firm offers a basic service that fails with probability 1—the special case of damaged goods (Deneckere and McAfee 1996).

This failure rate is strictly positive even though it is essentially free for the provider to offer a reliable service to all customers.

# 6 Damage Compensation

Instead of offering “service guarantees” that prevent failures, providers often use service recovery strategies to voluntarily deal with customer damages that result from service failure (Smith, Bolton and Wagner 1999).[18] This section shows how providers should design monetary compensation plans as a form of service recovery.

We assume that the provider specifies not only a price $p$ and a failure rate $q$, but also a compensation payment $k$ that is paid to the customer in the event that the service fails. Potential customers evaluate monetary outcomes $x$ with a (Bernoulli) utility function $u(x)$. A customer obtains utility $u(v-p)$ if the service succeeds, utility $u(v-d-p+k)$ if the service fails, and $u(0)$ from the outside option. We assume $u'(x) > 0$ and $u''(x) < 0$ to reflect that potential customers are risk averse.[19]

**Optimal Decisions.** The provider chooses the failure rate $q$, the price $p$ and the compensation payment $k$ in order to maximize the expected profit subject to the participation constraint of the potential customers:

$$\begin{aligned} \max_{q,p,k} \quad & \pi(q,p,k) = p - c(q) - qk \\ \text{s.t.} \quad & (1-q)\,u(v-p) + q\,u(v-d-p+k) \geq u(0). \end{aligned} \tag{5}$$

Optimal decision making in this setting can best be understood by looking at the allocation of risk via $p$ and $k$ for a given failure rate $q$. For $q = 0$, the compensation is irrelevant because it is never paid when the service is perfectly reliable, and it is optimal for the provider to set $p^* = v$. When $q = 1$, the compensation is always paid because the service is perfectly unreliable. In this case, the compensation corresponds to a price discount, and it is optimal to set the net price $p^* - k^*$ equal to $v - d$. At an interior solution

[18]There are also instances where providers are required by law to offer damage compensation. For example, EU and US regulations specify (minimum) compensation levels for flight delays and overbookings.

[19]The reason for introducing both extensions at once is that compensation payments are irrelevant as long as both parties are risk neutral. More specifically, both parties then care about $p$ and $k$ only via the expected net price $p - qk$, so that among the many optimal combinations, there always exists one involving $k = 0$.

for $q$, it is optimal for the provider to set $k^* = d$ and insure the risk-averse customers against potential damages. To fully extract the consumer surplus, the provider sets $p^* = v$. Finally, the optimal failure is determined exactly as in the base model. Thus, the following result holds:

**Proposition 5 (Compensation).** *Full damage compensation is optimal whenever $q^*$ is interior, that is, whenever $-c'(0) > d > -c'(1)$. For $q^* = 0$ and $q^* = 1$, compensation is irrelevant: it is never paid for $q^* = 0$ and amounts to a price discount for $q^* = 1$.*

Proposition 5 shows that the provider may offer an unreliable service even when potential customers are risk averse.[20] Intuitively, it is cheaper for the risk-neutral provider to deal with the consequences of service failures rather than solving the problem at the source and offering a reliable service. The result also shows that the provider should design the recovery strategy such that customers are fully insured against the consequences of service failures. The funds required to compensate against service failures are implicitly collected from the customers by charging them a higher price ($v$ instead of $v - q^*d$ in the base model).

# 7 Competitive Markets

Up to now, we have assumed that there is a single provider in the market. In this section, we relax this assumption to study how competitive forces affect optimal decisions about the failure rate and the price—and whether or not firms that offer imperfect service will be driven out of the market.

To capture a competitive environment, we consider two service providers, indexed by $i = 1, 2$, where each offers a single service to potential customers, each choosing a failure rate $q_i$ and a price $p_i$. Horizontal differentiation is à la Hotelling, with providers located at the extremes of the unit interval at $x_1 = 0$ and $x_2 = 1$, respectively. Vertical differentiation captures the notion that service failures reduce service quality. The unit cost of provider $i$ is given by $c_i(q_i)$, a specification that allows for cost heterogeneity among providers.

[20]Note that this result also holds if compensation is required by law. With damage litigation, expected damages become a cost for the provider, which implies that the efficient failure rate is unchanged because the total costs of failure to society at large are unaltered.

There is a unit measure of uniformly distributed potential customers, each with unit demand. The expected utility of the service purchased from provider $i$ is

$$u_i(q_i, p_i; x) = v - q_i d - \tau |x - x_i| - p_i,$$

where $v$ is the valuation for the service and $x \in [0,1]$ is the potential customer's preferred service offering. The parameter $\tau > 0$ measures the potential customers's sensitivity to horizontal mismatch $|x - x_i|$.[21] In this model of business stealing, the demand of service provider $i$ as a function of the failure rates $\mathbf{q} \equiv (q_1, q_2)$ and the prices $\mathbf{p} \equiv (p_1, p_2)$ is derived from the indifference condition $u_1(q_1, p_1; \hat{x}) = u_2(q_2, p_2; \hat{x})$ as

$$D_i(\mathbf{q}, \mathbf{p}) = \frac{1}{2\tau} \left( \tau - (q_i - q_j)d - (p_i - p_j) \right).$$

Hence, differences in demands among the competitors are driven by differences in failure rates and prices.

**Optimal Decisions.** Provider $i$ chooses the failure rate $q_i$ and the price $p_i$ in order to maximize the expected profit:

$$\max_{q_i, p_i} \quad \pi_i(\mathbf{q}, \mathbf{p}) = (p_i - c_i(q_i)) D_i(\mathbf{q}, \mathbf{p}).$$

In a Nash equilibrium, the optimal failure rate and price of provider $i$ are given by

$$q_i^c = c_i'^{-1}(-d) \quad \text{and} \quad p_i^c = \tau + \tfrac{1}{3}(q_j^c - q_i^c)d + \tfrac{1}{3}(2c_i(q_i^c) + c_j(q_j^c)).$$

The optimal failure is determined by the interplay of the cost structure and the customer damage—exactly as in the monopoly case in Section 2. The following result extends the "return on quality" approach to a competitive setting:

**Proposition 6 (Competition).** *It is optimal and economically efficient for provider i to offer a reliable service if $-c_i'(0) \leq d$ and to retain service failures if $-c_i'(0) > d$, $i = 1, 2$.*

Proposition 6 shows that competitive forces do neither drive a provider of unreliable service out of the market nor lower the optimal failure rate. The intuition for this counter intuitive result lies in a separability of each provider's optimization problem: the failure

[21] For further details, see Anderson, de Palma and Thisse (1992).

rate $q_i$ is optimally used to maximize the provider's margin per customer, whereas the optimal sales are implemented via choices of prices $p_i$. This result shows that the "return on quality" approach motivates providers in a competitive market environment to choose failure rates (and hence prices) that are optimal not only from their perspective but also from a societal point of view. Finally, as regards to prices, notice that for symmetric firms equilibrium prices simplify to $p_i^c = \tau + c(q^c)$, which illustrates firms' ability to raise price above marginal cost due to customer disutility from horizontal mismatch. More generally, for asymmetric firms, prices will differ, and it can be shown that the firm with a cost advantage captures a larger share of the market.

**Example.** To illustrate our results, assume that the cost functions take the quadratic form $c_i(q_i) = c_i(1-q_i)^2$, with $c_i > 0$, so that there is cost heterogeneity if $c_1 \neq c_2$. Then, in competitive equilibrium, the failure rate and price of provider $i$ are given by

$$q_i^c = 1 - \frac{d}{2c_i} \quad \text{and} \quad p_i^c = \tau + \frac{(4c_j - c_i)d^2}{12c_1c_2}.$$

These equilibrium quantities have intuitive properties: The failure rate increases in the cost parameter $c_i$ and decreases in the damage $d$, exactly as in the monopoly case. The equilibrium price decreases in the own cost parameter $c_i$ (because of the higher failure rate) and increases in the rival's cost parameter (because of the higher pricing power on the captive segment).

# 8 Conclusion

We showed that allowing for occasional service failures can be profitable and socially optimal for a service provider. When customers differ in perceived damages from a service failure, it can be profitable to use a hybrid strategy under which the provider randomizes service quality, but simultaneously offers a consistent service that is sold at a higher price than the unreliable service. This strategy, however, can create economic inefficiencies since the provider is motivated to retain service failures to capture surplus from customers with high damage costs who would pay a higher price for a reliable service. As part of such a hybrid strategy, the provider can invest in damage prevention strategies (such as backup and protection plans) or in less efficient monetary damage compensation (since the remuneration is offered after damage has occurred).

The analysis provides new insights into how a provider should design a service by simultaneously optimizing service failure and damage control. As damages incurred by service failure increase, the willingness to pay for an unreliable service decreases, so the provider has to lower the service price and vice versa. When the proportion of customers who perceive high damages increases, the provider can profit by deliberately increasing failure and damages, and then offer protection and compensation to those customers who would prefer to pay a higher price for reliable service.

Altogether, our results on optimal service design show how the key determinants of failures—costs, damages, and customer heterogeneity—drive the optimal failure rate and price. The results also provide a deeper understanding of how to design the optimal failure control. We show that—for both failure prevention and compensation—the decisions about the optimal failure rate and the design of the damage control are intertwined, and therefore should be made together by adopting a global view in the service-design process.

Beyond studying the profit impact, our analysis identifies conditions under which a profit orientation is detrimental to society because it leads to excessive service failures. This provides important new insights for managers who seek to engage in socially responsible service design by considering both the impact of their decisions on both customer well-being as well as on profit. Such an approach would also enable a transition to a so-called "solution economy" (Eggers and Macmillan 2013) that solves problems at the source rather than treating symptoms of service failures through protection strategies.

Future research could examine the impact of relaxing some of the model's key assumptions. On the demand side, it was assumed that potential customers know their expected damage and that their valuation of the service and the damage are additively separable in the expected utility function. Alternative assumptions would allow for a more general setup—for example, one could use a multiplicative functional relationship to study how this would impact the optimal failure rate and price. We also assumed that damages are perfectly correlated with the valuations of the service. This assumption can naturally be relaxed by assuming that there is an imperfect correlation instead. On the supply side, one could study how managerial biases about costs or customer damages would affect optimal failure rates.[22] Other assumptions that could be relaxed are that the unit costs, the customer damages, and the extent of customer heterogeneity are observable to the provider.

Regarding the technology, we have assumed that changes in the failure rate imply a smooth change in cost, whereas in reality there might be step costs or other technological constraints that prevent a smooth optimization. Incorporating more general assumptions would make the approach less tractable (if not intractable). In practice, failure rates are often re-optimized in the vicinity of current failures rates in response to changes in the business environment. In this case, smooth functions should provide a good enough approximation for the optimization of service-design decisions. Empirical studies could test the results implied from the analysis to understand if this assumption is a good approximation of the real world.

Future research could also study situations where the failure rate is influenced by the customer as well—a form of co-creation of service failures and the failure rate. Our insights help managers optimize the service process by allowing for occasional service failures and by applying different tools of damage control. Using prevention rather than compensation can reduce customer damages and enhance economic efficiency.

# Appendix

***Proof of Proposition 1.*** In order to fully extract the consumer surplus, the provider sets the optimal price for the service at $p^* = v - qd$. Substituting this back into the profit function (1) yields

$$\pi(q) = v - qd - c(q). \tag{A.1}$$

The necessary and sufficient Kuhn-Tucker conditions for a constrained maximum of (A.1) are:

$$\begin{aligned} -c'(q^*) &= d - \lambda_1 + \lambda_2, \\ \lambda_1 q^* = 0, &\quad \text{and} \quad \lambda_2(1 - q^*) = 0, \end{aligned} \tag{A.2}$$

where the $\lambda$s are nonnegative multipliers associated with the inequality constraints.

The provider sets $q^* = 0$ if $-c'(0) \leq d$: Since $\lambda_2 = 0$ , (A.2) implies that $-c'(0) = d - \lambda_1 \leq d$ (as $\lambda_1 \geq 0$ by assumption). Similarly, the provider sets $q^* = 1$ if $-c'(1) \geq d$: Since $\lambda_1 = 0$ , (A.2)

[22] In our model, if the provider underestimates the damages, this would result in lower (and socially inefficient) reliability and reduced sales (and profit) as the willing to pay is overestimated.

implies that $-c'(1) = d + \lambda_2 \geq d$ (as $\lambda_2 \geq 0$ by assumption). Consequently, the provider chooses $q^* \in (0,1)$ if $-c'(0) > d > -c'(1)$. Since the consumer surplus can be fully extracted, the profit of the provider coincides with welfare. Therefore, $q^*$ maximizes welfare and is economically efficient. □

***Proof of Lemma 1.*** At an interior solution, (A.2) boils down to $-c'(q^*) = d$. Hence, the optimal failure rate can be expressed as

$$q^* = c'^{-1}(-d).$$

Using the inverse function theorem,

$$\frac{\mathrm{d}q^*(d)}{\mathrm{d}d} = -\frac{1}{c''(q^*)} < 0.$$

Since $c''(q) > 0$, the optimal failure rate $q^*$ decreases in response to a higher $d$. It can readily be verified by inspection of Figure 3 that higher costs increase the optimal failure rate. □

***Proof of Proposition 2.*** As a preliminary step, we first establish the following claim from the main text concerning the relevance of constraints in the provider's optimization problem:

**Lemma A1.** *The provider's optimal choice of failure rates $q_i$ and prices $p_i$ must be such that* ($\mathrm{PC}_L$) *and* ($\mathrm{IC}_H$) *bind, whereas the remaining two constraints can be ignored.*

*Proof.* Notice first that the incentive constraints ($\mathrm{IC}_H$) and ($\mathrm{IC}_L$) jointly imply $(q_H - q_L)\Delta d \leq 0$, and thereby

$$q_H \leq q_L, \tag{A.3}$$

so that the failure rate targeted at the highs can never exceed that targeted at the lows. Now given our assumption that $\Delta v \geq \Delta d$ and given ($\mathrm{IC}_H$), ($\mathrm{PC}_H$) is satisfied whenever ($\mathrm{PC}_L$) is, implying that the highs buy whenever the lows do. We can therefore ignore ($\mathrm{PC}_H$). Moreover, ($\mathrm{PC}_L$) must bind in the optimum: otherwise, the provider can strictly increase profit (without violating any of the remaining constraints) by raising $p_L$ and $p_H$ by equally small amounts. But then ($\mathrm{IC}_H$) must bind, because otherwise profits can be raised by marginally raising $p_H$.

Hence, the problem reduces to maximizing $\pi(\mathbf{q}, \mathbf{p})$ subject to the binding versions of the constraints ($\mathrm{PC}_L$) and ($\mathrm{IC}_H$), and subject to ($\mathrm{IC}_L$). The latter in turn is equivalent to (A.3) (because ($\mathrm{IC}_H$) binds). Moreover, given our assumptions on the objective function, it is quickly seen that (A.3) can in fact be ignored: the unconstrained solution will always satisfy this constraint. □

This result implies that the optimal prices are given by $p_L = v_L - q_L d_L$ and $p_H = p_L + (q_L - q_H)d_H$. Substituting these prices back into the objective (3), the problem of choosing the segment-specific failure rates reduces to

$$\max_{q_L, q_H} \quad \pi(\mathbf{q}) = (1-\alpha)W_L(q_L) + \alpha(W_H(q_H) - \Delta v + q_L \Delta d), \tag{A.4}$$

where $W_i(q_i) \equiv v_i - q_i d_i - c(q_i)$ is the welfare generated by any type $i$ when purchasing the service that is targeted at their type. Additive separability of the objective in failure rates in turn implies that the failure rate offered to lows, $q_L$, must satisfy

$$\max_{q_L} (1-\alpha)W_L(q_L) + \alpha q_L \Delta d. \tag{A.5}$$

Following the same logic as in the proof of Proposition 1, the solution will involve $q_L > 0$ if the derivative of the objective function at zero is strictly positive, i.e. if $(1-\alpha)W_L'(0) + \alpha\Delta d > 0$. Using that $W_i'(q) = -d_i - c'(q)$ from the definition of $W_i(q)$ and rearranging then establishes that setting $q_L > 0$ is optimal if $d_L - \frac{\alpha}{1-\alpha}\Delta d < -c'(0)$. Using an argument completely analogous to that made in Section 2, this will be economically inefficient whenever $-c'(0) \leq d_L$. To complete the proof, note that setting $q_H = 0$ is optimal and economically efficient if $-c'(0) \leq d_H$, which automatically holds whenever $-c'(0) \leq d_L$.

As noted in the text (Footnote 14), this analysis assumes that the optimum involves serving both customer types with a differentiated offer. Alternatively, the provider might choose to serve only highs instead.[23] The following additional result shows that this is option is irrelevant (and hence the above analysis without loss of generality) so long as $\alpha$ is not too high, i.e. so long as there are enough lows:

**Lemma A2.** *Let $\alpha^S \equiv W_L(q_L)/W_H(q_L)$, where $q_L$ denotes the failure rate optimally offered to the lows when* both *types are served. Serving both types is optimal for $\alpha \leq \alpha^S$, whereas serving only $H$ types is optimal for $\alpha \geq \alpha^S$.*

*Proof.* The analysis of the case in which only highs are served is identical to that in Section 2, in that this will involve serving highs with the offer $(p_H, q_H)$ such that $p_H = v_H - q_H d_H$ (binding (PC$_H$)) and such that $q_H$ maximizes $W_H(q_H)$ (the same $q_H$ as in the above case where both types are served), which results in expected profits the size of $\alpha W_H(q_H^*)$. In contrast, when serving both types, the optimal offer derived above yields profits $(1-\alpha)W_L(q_L) + \alpha(W_H(q_H^*) - \Delta v + q_L \Delta d)$

[23] A third option would be not to serve any types at all. As previously in the case of homogenous customers, we will ignore that option. It is easy to derive parameter restrictions such that this involves no loss of generality.

(see (A.4)). The claim then immediately follows from comparing the two profits and simplifying (using that $\Delta v - q\Delta d = \Delta W(q) \equiv W_H(q) - W_L(q)$). □

The conditions under which both types are offered the same (pooling) offer are standard and therefore omitted. □

***Proof of Lemma 2***. At an interior solution, the necessary and sufficient first-order condition for a maximum of (A.5) is given by

$$(1-\alpha)[-d_L - c'(\tilde{q}_L)] + \alpha\Delta d = 0.$$

Hence, the optimal failure rate for lows can be expressed as

$$\tilde{q}_L = c'^{-1}\left(-d_L + \frac{\alpha\Delta d}{1-\alpha}\right). \tag{A.6}$$

Since $\alpha$ and $\Delta d$ are strictly positive, $c'(q) < 0$ implies that $\tilde{q}_L > q^*$. Inspection of (A.6) reveals that the extent of the distortion increases in $\alpha$ and $\Delta d$. Finally, note that the information rent for a high is the difference between $p_H^* = v_H$ and $\tilde{p}_H = v_L + \tilde{q}_L\Delta d$, given by $\Delta v - \tilde{q}_L\Delta d$. This implies that the money left on the table for highs is given by $\alpha(\Delta v - \tilde{q}_L\Delta d)$, as can be seen in (A.4). □

***Proof of Proposition 3***. Using backups is profitable for the provider if it increases profit over the level attained in the benchmark case. Specifically, this requires that

$$v - c(q) - qw > v - c(q) - qd.$$

Therefore, if $w < d$, the provider markets the service using backups and chooses $q_b$ in order to maximize $v - c(q_b) - q_b w$, which is also economically efficient. (Note that the profit coincides with welfare as the consumer surplus is fully extracted.) Instead, if $w \geq d$, the provider offers the service without procuring backups and chooses $q$ in order to solve (1). Using Proposition 1, it immediately follows that $q_b^* > q^*$ if $w < d$. □

***Proof of Proposition 4***. Suppose that the cost function $c(q)$ satisfies the property $-c'(0) \equiv \varepsilon$, where $\varepsilon > 0$ is a small but positive number. Since the profit function (4) is equivalent to the profit function (3), we can apply Proposition 2: If

$$-c'(1) < d_L - \frac{\alpha\Delta d}{1-\alpha} < \varepsilon \leq d_L,$$

then the basic service fails with probability $q \in (0,1)$ and is priced at $p = v_L - q_L d_L$ (see proof of Lemma 2). The protection add-on is priced at $f = qd_H$. Service differentiation is profitable but economically inefficient as $\varepsilon \leq d_L$. □

***Proof of Proposition 5***. The necessary and sufficient Lagrange conditions concerning optimal choice of $p$ and $k$ in optimization problem (5) are:

$$1-\lambda\left[(1-q)u'(v-p)+qu'(v-d-p+k)\right]=0$$
$$-q+\lambda qu'(v-d-p+k)=0.$$

Combining to eliminate $\lambda$ (and using that $q\neq 0$ by presumption) yields $u'(v-p)=u'(v-d-p+k)$. Since $u''<0$, this immediately implies equality of the arguments, and hence $k=d$. This establishes that, whatever the (optimal) choice of $q$, the customer receives full damage compensation. Given this, the potential customer faces no uncertainty, so that the remaining analysis proceeds exactly as in the case above without compensation and with risk neutral customers, leading to the same failure rate as derived there. Formally, inserting $k=d$ and $p=v$, the remaining optimization problem of the provider is exactly as above. Thus, the same conditions for an interior solution for $q$ apply. $\square$

***Proof of Proposition 6***. In equilibrium, each provider $i$'s choice of $q_i$ and $p_i$ must solve

$$\max_{q_i,p_i} \quad \pi_i(\mathbf{q},\mathbf{p})=(p_i-c_i(q_i))D_i(\mathbf{q},\mathbf{p}), \tag{A.7}$$

where $D_i(\mathbf{q},\mathbf{p})$ is provider $i$'s demand.

To see the role of $q_i$ in this problem, consider a transformed version of the problem, where, rather than choosing $q_i$ and $p_i$, each provider chooses $q_i$ and $\hat{v}_i\equiv v-q_id-p_i$ (the latter might be thought of as the expected utility which the service offers to customers, net of price and expected damage, but gross of the costs due to horizontal mismatch). To every $(q_i,p_i)$-pair, there corresponds a unique $(q_i,\hat{v}_i)$-pair, and vice versa, making this a proper change of variables. Now utility obtained by a customer located at $x$ from service $i$ can be written as $u_i(\hat{v}_i;x)=\hat{v}_i-\tau\,|x-x_i|$. Importantly, fixing $\hat{v}_i$, customers' utilities are independent of $q_i$, and so is demand $D_i$. Using this and that $p_i=v-q_id-\hat{v}_i$ from our change of variables, problem (A.7) can be reformulated as

$$\max_{q_i,\hat{v}_i} \quad \pi_i(\mathbf{q},\hat{\mathbf{v}})=(v-q_id-\hat{v}_i-c_i(q_i))D_i(\hat{v}_i,\hat{v}_j), \tag{A.8}$$

from which it immediately follows that $q_i^c$ must maximize the first factor, $v-q_id-\hat{v}_i-c_i(q_i)$, which corresponds to the provider's margin per customer.

Having thus determined $q_i^c$, we can now return to our original, untransformed problem in (A.7), where prices are now characterized by the two first-order conditions on prices alone:

$$\frac{\partial\pi_i}{\partial p_i}=D_i(\mathbf{q}^c,\mathbf{p})+(p_i-c_i(q_i^c))\frac{\partial D_i(\mathbf{q}^c,\mathbf{p})}{\partial p_i}=0. \tag{A.9}$$

Assuming an equilibrium in which all potential customers are served and both firms sell positive amounts, equilibrium demand is characterized by a customer $\hat{x}$ who is indifferent between buying from either firm, such that $D_1(\mathbf{q},\mathbf{p}) = \hat{x}$ and $D_2(\mathbf{q},\mathbf{p}) = 1-\hat{x}$. Solving the indifference condition $u_1(q_,p_1;\hat{x}) = u_2(q_2,p_2;\hat{x})$ for $\hat{x}$ yields

$$\hat{x} \equiv D_1(\mathbf{q},\mathbf{p}) = \frac{1}{2\tau}\left(\tau - (q_1-q_2)d - (p_1-p_2)\right).$$

Using this in (A.9) and solving yields equilibrium prices

$$p_i = \tau + \tfrac{1}{3}d(q_j^c - q_i^c) + \tfrac{1}{3}(2c_i(q_i^c) + c_j(q_j^c)),$$

as claimed.

It remains to be shown that a social planner who chooses $q_i$ and $p_i$ so as to maximize welfare would choose the same failure rates $q_i^c$. To see this, notice that such a planner's problem can be formulated as

$$\max_{\mathbf{q},\mathbf{p}} \sum_{i=1}^{2}(v - dq_i - c_i(q_i))D_i(\mathbf{q},\mathbf{p}) + \int_0^{D_1(\mathbf{q},\mathbf{p})} x\tau dx + \int_{D_1(\mathbf{q},\mathbf{p})}^{1}(1-x)\tau dx.$$

Using the above change of variables to substitute $\hat{v}_i$ for $p_i$, this can be rewritten as

$$\max_{\mathbf{q},\hat{\mathbf{v}}} \sum_{i=1}^{2}(v - dq_i - c_i(q_i))D_i(\hat{\mathbf{v}}) + \int_0^{D_1(\hat{\mathbf{v}})} x\tau dx + \int_{D_1(\hat{\mathbf{v}})}^{1}(1-x)\tau dx,$$

which again immediately shows that $q_i$ must maximize a provider's margin per customer, as claimed.[24] □

# References


Anderson, S.P., De Palma, A. and Thisse, J.F., 1992. *Discrete choice theory of product differentiation*. MIT press.

Biyalogorsky, E. and Gerstner, E., 2004. Contingent pricing to reduce price risks. *Marketing Science*, 23(1), pp. 14–155.

Bloomberg (July 12). United wants to sell your seat to someone else for more money. https://www.bloomberg.com/news/articles/2017-07-12/united-wants-to-sell-your-seat-to-someone-else-for-more-money


[24] A similar argument also shows that a monopolist who operates both service outlets will likewise choose $q_i^c$. This again shows that rivalry in the market does not lower the failure rates.

Bulow, J., 1986. An economic theory of planned obsolescence. *The Quarterly Journal of Economics*, 101(4), pp. 729–749.

Chase, R.B. and Stewart, D.M., 1994. Make your service fail-safe. *Sloan Management Review*, 35(3), pp. 35–44.

Chen, R.R., Gerstner, E. and Yang, Y., 2009. Research Note—Should Captive Sardines Be Compensated? Serving Customers in a Confined Zone. *Marketing Science*, 28(3), pp. 599–608.

Chen, R.R., Gerstner, E. and Yang, Y., 2012. Customer bill of rights under no-fault service failure: Confinement and compensation. *Marketing Science*, 31(1), pp. 157–171.

Chen Y., Koenigsberg, O. and Zhang J., 2016. Pay-as-you-wish pricing. *Marketing Science*, in press.

Deneckere, R.J. and Preston McAfee, R., 1996. Damaged goods. *Journal of Economics & Management Strategy*, 5(2), pp. 149–174.

Dukes, A.J. and Zhu, Y., 2016. Why customer service frustrates consumers: Exploiting hassle costs by a tiered customer service organization. Working paper, USC Marshall School of Business, available at https://ssrn.com/abstract=2832927.

Eggers, W.D., and Macmillan, P., 2013. The Solution Revolution: How Business, Government, and Social Enterprises are Teaming Up to Solve Society's Toughest Problems. Harvard Business Review Press.

Fay, S. and Xie, J., 2008. Probabilistic goods: A creative way of selling products and services. *Marketing Science*, 27(4), pp. 674–690.

Gallego, G. and Stefanescu, C., 2012. Services engineering: Design and pricing of service features. In *The Oxford Handbook of Pricing Management*.

Gerstner, E. and Libai, B., 2006. Invited Commentary—Why Does Poor Service Prevail? *Marketing Science*, 25(6), pp. 601–603.

Hogreve, J. and Gremler, D.D., 2009. Twenty years of service guarantee research: A synthesis. *Journal of Service Research*, 11(4), pp. 322–342.

Levitt, T., 1972. Production-line approach to service. *Harvard Business Review*, 50(5), pp. 41–52.

Ma, L., Sun, B. and Kekre, S., 2015. The Squeaky Wheel Gets the Grease—An empirical analysis of customer voice and firm intervention on Twitter. *Marketing Science*, 34(5), pp. 627–645.

Moorthy, S. K. and Png, I.P.L., 1992. Market segmentation, cannibalization, and the timing of product introductions. *Marketing Science*, 38(3), pp. 345–359.

Nelson, P., 1974. Advertising as information. *Journal of Political Economy*, 82(4), pp. 729–754.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L., 1985. A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), pp. 41–50.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L., 1988. Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), pp. 12–40.

Rust, R.T. and Zahorik, A.J., 1993. Customer satisfaction, customer retention, and market share. *Journal of Retailing*, 69(2), pp. 193–215.

Rust, R.T., Zahorik, A.J. and Keiningham, T.L., 1995. Return on quality (ROQ): Making service quality financially accountable. *Journal of Marketing*,

Rust, R.T., Moorman, C. and Dickson, P.R., 2002. Getting return on quality: Revenue expansion, cost reduction, or both? *Journal of Marketing*, 66(4), pp. 7–24.

Rust, R.T. and Chung, T.S., 2006. Marketing models of service and relationships. *Marketing Science*, 25(6), pp. 560–580.

Rust, R.T. and Huang, M.H., 2012. Optimizing service productivity. *Journal of Marketing*, 76(2), pp. 47–66.

Schmidt, K.M., Spann, M. and Zeithammer, R., 2014. Pay what you want as a marketing strategy in monopolistic and competitive markets. *Management Science*, 61(6), pp. 1217–1236.

Shapiro, C. and Varian, H.R., 1998. Versioning: The smart way to sell information. *Harvard Business Review*, pp. 106–114.

Shin, J. and Sudhir, K., 2010. A customer management dilemma: When is it profitable to reward one's own customers? *Marketing Science*, 29(4), pp. 671–689.

Smith, A.K., Bolton, R.N. and Wagner, J., 1999. A model of customer satisfaction with service encounters involving failure and recovery. *Journal of Marketing Research*, 36(3), pp. 356–372.

Varian, H.R., 1980. A model of sales. *The American Economic Review*, 70(4), pp. 651–659.

Waldman, M., 1993. A new perspective on planned obsolescence. *The Quarterly Journal of Economics*, 108(1), pp. 273–283.

Wu, T., (2014). Why Airlines Want to Make You Suffer. *The New Yorker*, December 26.

Zhang, Z., Joseph, K. and Subramaniam, R., 2015. Probabilistic selling in quality-differentiated markets. *Management Science*, 61(8), pp. 1959–1977.