Electronic Thesis and Dissertation Repository

6-29-2021 10:30 AM

Trust and Bias in Peer-to-Peer Ratings: Why Peer-to-Peer Service Ratings are Nearly Always Positive, and How They Can be Fixed

Michael Moorhouse, *The University of Western Ontario*

Supervisor: Cotte, June, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Business
© Michael Moorhouse 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Marketing Commons

Recommended Citation

Abstract

Transactions in the peer-to-peer sharing economy carry high risk and uncertainty. Consumers exchange with non-professional providers with whom they have no past history, and must rely on ratings and reviews for choice selection. However, there is a large positive bias in the ratings, making differentiation difficult, and causing some consumers to lose trust. Despite these concerns, little progress has been made to demonstrate the cause of the bias or how it can be fixed. I address this gap by demonstrating that consumers evaluate peer-peer experiences based on trust. This trust evaluation, in concert with network and social factors, contributes to the bias.

Research on service evaluation is often informed by the expectancy disconfirmation process (Oliver, 1980, 2010). Consumers compare a provider's performance against prior expectations; the resultant satisfaction or dissatisfaction leads to online ratings. I demonstrate that the process works differently for peer-to-peer services; a consumer's determination of whether a provider met expectations has an effect on ratings beyond the effect of satisfaction (Study 1). When uncertainty and risk are high, a provider demonstrates that they can be trusted by meeting a consumer's prior expectations (Study 2). Contextual factors in peer-to-peer networks cause consumers to feel that their ratings are more important to peer providers, and that they may need to justify ratings. This elevates trust as an important driver of ratings at the expense of satisfaction, because satisfaction is more subjective and more difficult to justify (Study 3).

Consumers may give peer providers positive ratings even if performance is worse than expected. Standards of evaluation are relatively unclear for peer-to-peer services (making it more difficult to identify performance failure), and social norms of gratitude and empathy motivate consumers to forgive peer providers for unreliable service (Studies 4 and 5). Negative ratings for peer providers may result only if consumers believe that a provider caused and controlled a negative outcome, which suggests a lack of integrity (Study 6). I demonstrate that platforms can attenuate the positive bias by making ratings anonymous, by clearly defining service standards, and by increasing perceived controllability by providers for expectations and performance failure.

Summary for Lay Audience

The peer-to-peer sharing economy is growing quickly behind platforms such as Airbnb and Uber that help people rent or share their skills and belongings with other consumers. Online ratings and reviews are extremely important for consumers of peer-to-peer services because they establish trust with unknown (and mostly non-professional) providers. However, nearly all peer-to-peer ratings are five-stars, which makes it difficult for consumers to distinguish between providers. It suggests that peer-to-peer ratings may be biased, and may not reflect a provider's true quality. I attempt to determine the cause of this positive ratings bias, and provide solutions to fix the bias.

The dissertation is comprised of six studies. I first explore how consumers of peer-to-peer services evaluate their experiences differently than consumers of traditional services. Research shows that for consumers who rent from a traditional business, their satisfaction is the main driver of the ratings decision. I show that this is not true for peer-to-peer services (Study 1). In peer-to-peer services, consumers experience higher feelings of risk and uncertainty because they are dealing with strangers. I show that when risk and uncertainty are high, a provider who meets a consumer's expectations demonstrates their trustworthiness (Study 2).

Next, I demonstrate that the feeling of trust in the provider is directly reflected in peer-to-peer ratings and may lead to positive ratings even when performance is worse than expected. This is because peer-to-peer services have important differences compared to

traditional services that cause peer-to-peer consumers to feel that they need to justify their ratings decisions, and to feel gratitude and empathy toward peer providers (Studies 3-5). This leads to high ratings even if a consumer is relatively unsatisfied, as long as the provider was relatively trustworthy. I show that peer-to-peer consumers give low ratings only if they feel that an untrustworthy provider caused and could have prevented a service failure (Study 6), but that it is difficult for peer-to-peer consumers to make these assessments. Recognizing this, I test three ways that platforms can reduce the ratings bias by reducing the perceived need to justify ratings and by making it easier to recognize service failures.

Acknowledgments

I would like to extend my sincere gratitude to all of the individuals who helped me to achieve this milestone. First, I thank my supervisor June Cotte for her friendship, and for her guidance for my thesis, my job search, and my career. June has been unfailingly supportive, and I could not have imagined a better fit for me in choosing an advisor, mentor, and friend for this experience. I would also like to thank my supervisory committee, Miranda Goode and Matt Thomson, for their support, for continuing to push me, and for opening my eyes to new ideas.

I would like to thank the rest of the Ivey Marketing faculty. I've learned a lot from each of them about how to be a better researcher, teacher, and colleague. I would like to thank especially Kersi Antia, Xin Wang, Mike Taylor, Colin McDougall, and John White. I also thank my fellow doctoral students, especially Fernando Naranjo and his family for their continued friendship, advice, and moral support.

Finally, I thank my family. Thank you to my children, Cohen and Christina-Rose, for constantly putting a smile on my face, and for reminding me about what's most important in my life. My biggest thanks goes to my wife, Melissa. Without her support, this would not have been possible. I thank her for always believing in me, and for agreeing to take this journey together. I thank her for all of her sacrifices, and for supporting me to leave a successful career to pursue my dream. I celebrate and share this achievement with her.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1   Introduction

This research is inspired by my experience as a consumer of peer-to-peer services in the sharing economy. I have a background in brand management, and am fascinated by how peer providers can attempt to demonstrate their quality and trustworthiness to potential customers without the benefit of traditional signals (e.g., brand names, logos, or seals-of-approval; Schurr & Ozanne, 1985). Ratings and reviews are the primary means of provider differentiation in the sharing economy but, as demonstrated in the following anecdote, a positive bias in peer-to-peer ratings makes ex ante evaluation difficult and could lead to a dissolution of trust.

Dave is heading to another city for a one-night stopover. Rather than rent a hotel room, he'd rather just rent a room in someone's home through AirBnB. He has heard this is a good option, but worries about sharing space with a stranger. When he goes on the site, he notices that most (if not all) of the hosts are rated five stars, which makes him wonder how useful the reviews are. He just isn't sure that he can trust a stranger to host him.

The peer-to-peer sharing economy has fundamentally changed how many people consume products and services. Sharing economy platforms like Airbnb and Uber have experienced tremendous growth, and the industry now represents a significant share of the global economy (Caldieraro et al., 2018). Through networked technology, platforms

facilitate exchanges between consumers and non-professional (i.e., peer) providers who share or rent "underutilized assets from spaces to skills to stuff" (Botsman, 2013). Consumers gain temporary access to a provider's goods and services, which has implications for risk, reputation, and trust (Eckhardt et al., 2019). The most important factor for the continued growth of the sharing economy is trust (Cheng, 2016; Ter Huurne et al., 2017), because of the relatively high levels of risk and uncertainty. Peer-to-peer exchanges take place between strangers (Schor, 2014), often with little institutional controls such as formal screening or guarantees (Belk, 2013). Exchanges that begin online may proceed into private homes and vehicles, contributing to performance, property, and safety risks.

For consumers who don't have the benefit of past experience with a provider, the online review system is the critical source of reputation and trust. To be effective, consumers must perceive the feedback system to be *unbiased* (i.e., it provides an honest, accurate reflection of past behavior) and *diagnostic* (i.e., it is helpful for differentiating between providers). However, despite a high level of heterogeneity in provider quality, the overwhelming majority of peer-to-peer reviews and ratings are positive. For example, 94% of Airbnb properties were rated at 4.5 stars or higher (Zervas et al., 2015), and 90% of UberX trips were given 5 stars (Athey et al., 2018). This extreme distribution suggests that peer-to-peer ratings may be biased because they do not reflect the true quality differences between providers (Bridges & Vásquez, 2016). The likely result is that very high average ratings may cause consumers to become skeptical of feedback accuracy (Chevalier & Mayzlin, 2006). Diagnosticity also suffers, because low variability in

ratings makes it difficult to distinguish quality differences. This can erode trust, and consumers may respond to poor quality experiences by leaving the platform (Nosko & Tadelis, 2015). More research is critically needed to explain the positive bias in peer-to-peer ratings and to determine how to fix it. The goal of this dissertation is to illustrate how the differences between peer-to-peer and commercial services affect how consumers evaluate peer-to-peer experiences, and to show why this leads to a positive bias in ratings. I then propose and test different solutions to fix the bias and reduce the average rating.

With the six studies included in this dissertation, I demonstrate that the performance evaluation process works differently for peer-to-peer services than for services delivered by commercial businesses. The traditional view of performance evaluation is the expectancy disconfirmation model (Oliver, 1980). The model proposes that a consumer evaluates a service experience against their prior expectations. This leads to feelings of satisfaction or dissatisfaction, and those feelings affect the ratings decision. However, I show that for peer-to-peer services, a consumer's evaluation of whether the provider met expectations affects ratings beyond the impact of satisfaction (*Study 1*). Specifically, I demonstrate that expectancy disconfirmation can lead to perceptions of trust. The inherent uncertainty and risk in peer-to-peer sharing increase the importance of meeting expectations. Providers that meet expectations in an uncertain and risky environment demonstrate that they can be trusted, and this trust is reflected in a consumer's rating of the provider (*Study 2*). Understanding that peer-to-peer ratings are affected by trust (in addition to satisfaction), I next show how this evaluation can lead to highly positive ratings when the consumer's expectations are *merely met* (rather than exceeded), and

even sometimes when performance is worse than expected. The expectancy

disconfirmation model predicts that when expectations are merely met (rather than

exceeded), satisfaction will be moderate, and should lead to moderate ratings. However, I

demonstrate that meeting expectations in peer-to-peer sharing (versus commercial

businesses) can lead to highly positive ratings for two reasons: 1) ratings are more

important to peer providers than traditional providers, and 2) relatedly, consumers feel

that they need to justify the ratings they give providers. This increases the importance of

trust, and reduces the importance of satisfaction which is more subjective and thus more

difficult to justify. Platforms that help consumers feel that their rating is less important to

providers can partially attenuate the positive ratings bias *(Study 3)*.

The expectancy disconfirmation model predicts that when performance is *worse* than

expected, consumers will be dissatisfied, and this should lead to low ratings. However, I

show that negative disconfirmation in peer-to peer services can lead to positive ratings

because gratitude and empathy encourage consumers to forgive peer providers when

expectations are missed, especially when standards of evaluation are relatively unclear, as

they often are in peer-to-peer services (*Studies 4 and 5*). Thus, platforms that set clear

standards for peer-to-peer services can help to attenuate the positive ratings bias.

Evaluations of trust include assessments of a provider's reliability and integrity.

Consumers may forgive peer providers for service that is unreliable, but missed

expectations can lead to negative ratings if consumers infer that a provider's performance

failure was controllable. Controllability implies intentionality, which leads consumers to

question the provider's integrity. With this in mind, in *Study 6* I test a novel ratings system that can attenuate the positive ratings bias by asking providers to rate themselves on key attributes. Provider-authored ratings act as implicit commitments, and consumers should perceive that missed expectations based on those commitments are controllable. Thus, providers who rate themselves higher than their true performance level will be considered dishonest, and consumers should feel justified to assess them negatively. The conceptual models for the set of six studies are in Figure 1.

2   Conceptual Development

Peer-to-peer platforms need to address positive ratings bias, because accurate ratings are critical for establishing trust and preventing customer defection. Possible explanations for the bias include concern for managing one's public image to other members of the sharing network (Mulshine, 2015), pressure to not harm a peer provider's ability to earn income (Filippas et al., 2018), and socially induced reciprocity based on personal connections (Fradkin et al., 2015). Each of these may have merit, but none have been experimentally tested, nor positioned within a framework that explains how they affect the ratings decision process. I address this by exploring the differences in the expectancy disconfirmation process for peer-to-peer services, and how these differences contribute to biased ratings. In doing so, I suggest and test possible solutions for how sharing platforms can fix the problem. Next, I review the established literature to develop the propositions that will be tested in the six studies. A summary of the propositions are in Appendix A.

Panel A. Proposed Effect of Trust and Satisfaction on Peer-to-Peer Ratings
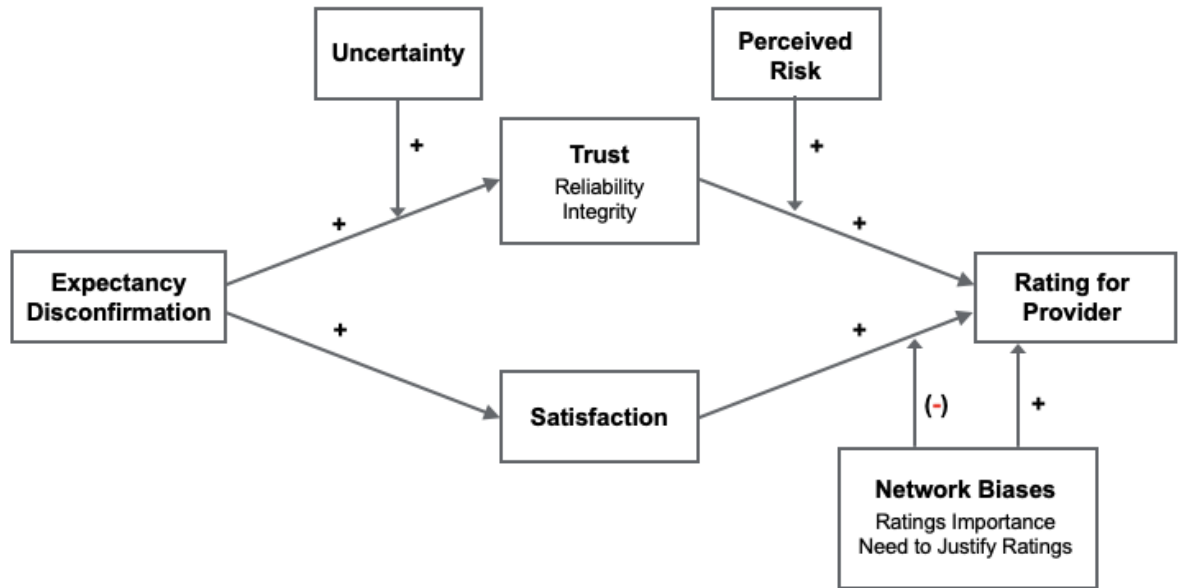
Panel B. Negative Disconfirmation and Positive Ratings Bias: Proposed Differential Effect of Perceived Reliability and Integrity on Peer-to-Peer Ratings
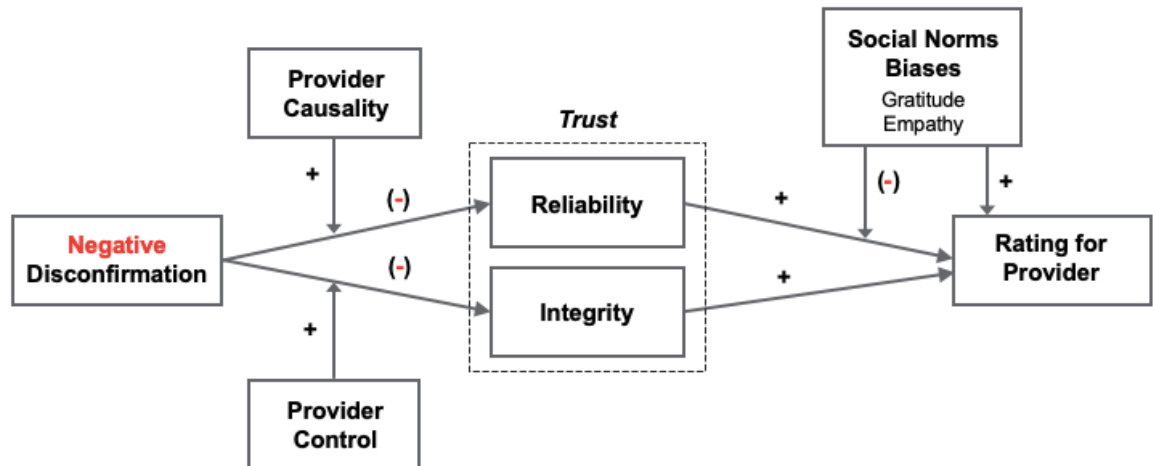
Figure 1: Proposed Conceptual Models

2.1   Product Quality and Electronic Word-of-Mouth

Consumers often rely on opinions from other consumers to help them make purchase decisions. Word-of-mouth is valuable because it is considered to be less biased than commercial messages (Godes & Mayzlin, 2004; Hamilton et al., 2014). The Internet has greatly expanded word-of-mouth communication options (Dellarocas, 2003), including through descriptive product reviews, ratings, upvotes and downvotes, social media and discussion forums, and many others. Collectively, electronic word-of-mouth is defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Hennig-Thurau et al., 2004, p. 39).

When making purchase decisions in online platforms, consumers place the most importance on the aggregate product rating (De Langhe et al., 2016). Aggregate ratings (i.e., the mean score for all of the individual ratings for a product) have been shown to drive online sales for products across many categories such as books (Chevalier & Mayzlin, 2006), toiletries (Moe & Trusov, 2011), and video games (Zhu & Zhang, 2010). This is because ratings are widely believed (by consumers and researchers alike) to be an important indicator of quality (Engler et al., 2015; Hu et al., 2017; Simonson & Rosen, 2014). Indeed, aggregate ratings, which are often assessed on a 5-star scale, have a stronger influence on consumers' quality inferences than other implicit online quality cues such as the number of ratings or the price of the product (De Langhe et al., 2016).

However, researchers have begun to question the accuracy of ratings because they often

correlate poorly with objective quality measures such as *Consumer Reports* scores (De Langhe et al., 2016). This may be due to biases in either the individual rating from an individual consumer, or in the aggregate distribution. For individual ratings, social influence and impression management concerns may cause a consumer to post a rating that is higher or lower than their true opinion. For example, negative opinions from other consumers (in the form of prior ratings) can cause a consumer to post a negative rating so that they don't appear indiscriminate (Schlosser, 2005). Positive prior ratings can also influence ratings. A consumer that has a positive product experience may post a negative rating to affirm their distinctiveness, while a consumer that has a negative experience may post a positive rating because they feel pressure to conform to popular opinion (Sridhar & Srinivasan, 2012). These herding and differentiation effects are common in many product and service review platforms (Lee et al., 2015).

The difference between ratings scores and objective quality may also be due to biases in the total distribution of ratings. Rather than a normal distribution, online ratings for most commercial products and services cluster in the extremes, with many positive ratings, a moderate amount of negative ratings, and very few neutral opinions. This "j-shaped" ratings distribution may be due to self-selection bias (Hu et al., 2017; Moe & Schweidel, 2011). Consumers with extremely negative or extremely positive feelings about a product are much more likely to leave an online review. To understand this bias, we must consider consumers' motivation for providing word-of-mouth, which is most often explained as a consequence of satisfaction (Brown et al., 2005).

2.2   Expectancy Disconfirmation, Satisfaction, and Ratings

The path from product experience to satisfaction to electronic word-of-mouth can be

explicated through the well-established expectancy disconfirmation model (Oliver, 1980,

2010). The model proposes that, prior to purchase, consumers form initial expectations

about how the product or service will perform. These expectations may come from prior

experience with the product or category, from claims made by the marketer, or from

opinions expressed by other consumers (Woodruff et al., 1983). When a consumer

experiences the product, they compare actual performance (i.e., perceived quality) with

expected performance. If the product or service performs better than expected, positive

disconfirmation occurs, and consumers are satisfied. If the product or service performs

worse than expected, negative disconfirmation occurs, and consumers are dissatisfied.

In the context of the dissertation, I take the view that satisfaction is a post-experience

response to the evaluation of a specific purchase occasion (i.e., "transaction-specific"

satisfaction; e.g., Halstead et al., 1994; Oliver, 1993; Spreng et al., 1996) which is the

dominant view from the expectancy disconfirmation model. It is important to note that

this view is different from "cumulative satisfaction" which is an overall evaluation based

on a product's past, current, and future performance, and which is often used in the

service quality and relational exchange literatures (Anderson et al., 1994; Boulding et al.,

1993; Fornell, 1992). I measure transaction-specific satisfaction because most sharing

economy consumers have no past history with their provider, and exchanges are likely to

be one-time only.

Transaction-specific satisfaction has been found to influence a variety of purchase-related outcomes including willingness-to-pay, repeat purchase, loyalty intentions, and word-of-mouth (Agustin & Singh, 2005; Anderson & Sullivan, 1993; Eisenbeiss et al., 2014; Meuter et al., 2000). This research shows that if the satisfaction response to the expectancy disconfirmation evaluation is strong enough, it motivates consumers to take action (Anderson, 1998; Eisenbeiss et al., 2014). Specifically, satisfaction motivates word-of-mouth as a means to persuade others, or in the case of extreme dissatisfaction, as a way to vent or harm a provider (Berger, 2014).

Posting online reviews and ratings takes effort, and often requires the consumer to attach their name (or username) to their review. Thus, many consumers choose not to leave a review (Avery et al., 1999). Consumers who experience moderate satisfaction are especially unlikely to expend the effort (Schoenmüller et al., 2018) because the satisfaction response is not strong enough to motivate action. Therefore, their moderate opinions are often not reflected in aggregate ratings. Thus, the difference between aggregate ratings and objective quality measures can partly be explained by self-selection bias that excludes many moderate opinions from the total distribution of ratings.

The self-selection problem demonstrates that satisfaction and dissatisfaction influence not only *whether* a consumer decides to perform word-of-mouth (for example, posting an online rating), but also *what* the consumer posts, in terms of the rating itself. Highly satisfied consumers post highly positive ratings (5-stars), and highly dissatisfied consumers post highly negative ratings (1-star). This allows them to either reward or

punish providers. Recognizing this, some researchers have begun to model ratings not as an outcome of quality, but rather as an outcome of satisfaction (e.g., Engler et al., 2015; Moe & Schweidel, 2011). These models measure both product performance and initial expectations, or a comparison between the two (expectancy disconfirmation), and treat expectancy disconfirmation as a proxy for satisfaction. The models therefore make an implicit assumption that the full effect of expectancy disconfirmation on online ratings is mediated through feelings of satisfaction or dissatisfaction. As we will see, this may not be accurate, especially for peer-to-peer services.

Similar to commercial services, peer-to-peer platforms provide consumers with an opportunity to express electronic word-of-mouth. In fact, reviews and ratings from other consumers are especially important for peer-to-peer services because providers are unknown to consumers, and there are few other sources of online reputation or quality assurance. Like most commercial services, most peer-to-peer platforms use a 5-star ratings scale. But peer-to-peer ratings are skewed even more to the positive, and have even less variance than commercial ratings (Bolton et al., 2012; Chevalier & Mayzlin, 2006). Rather than a "j-shaped" distribution, the ratings for peer-to-peer services are nearly all positive. In addition to the previous examples for Airbnb and Uber, 91% of oDesk services were rated at 4 or 5 stars (Horton & Golden, 2015), and 98% of BlaBlaCar trips were rated 5 stars (Slee, 2013).

Can this extreme positive ratings distribution be explained by satisfaction and/or self-selection? Evidence suggests that this may not be the case (see Table 1 for a summary).

Because satisfaction results from the evaluation of expectancy disconfirmation, satisfaction could explain the extreme positive skew in peer-to-peer ratings only if consumers of peer-to-peer services were significantly more likely to have their expectations exceeded than commercial consumers. This would result from either significantly higher perceived quality in peer-to-peer services, and/or significantly lower expectations. Peer-to-peer services are often delivered by non-professional providers. It is therefore unlikely that they will deliver better performance and higher quality than commercial providers on average. Further, the varying levels of commitment, experience, and expertise by peer providers makes peer provider quality more heterogenous than commercial services (Fradkin et al., 2015). Thus it is unlikely that quality is consistently high in peer-to-peer services. Expectations are also not likely to be the main driver of the positive bias in peer-to-peer ratings. Although there is some evidence that peer-to-peer consumers may have slightly lower expectations, this does not fully explain the ratings distribution (Fradkin, 2017). Indeed, consumers purchase the products and services that best balance the trade-off between costs and expected quality. On balance, the majority of consumers (both in commercial and peer-to-peer platforms) should have relatively high expectations and a positive initial disposition toward the product, or else they would not have purchased that product (Schoenmüller et al., 2018).

Table 1: Evaluation of Possible Causes for Consistently High Peer-to-Peer Ratings

Observations	Implications for Peer-to-Peer *Experiences*	Implications for Peer-to-Peer *Ratings*
Peer providers are mostly not professional	The quality of peer providers is unlikely to be significantly higher than commercial providers	Consistently high quality is not the cause of positive bias in peer-to-peer ratings
There is high variance in peer providers' experience, expertise, commitment, and quality of shared assets	Peer providers' quality is more heterogeneous than commercial providers	Consistently high quality is not the cause of positive bias in peer-to-peer ratings
In general, consumers choose products in which they have relatively high expectations	Initial expectations are unlikely to be significantly lower in peer-to-peer services	Consistently low expectations are not the cause of positive bias in peer-to-peer ratings
Peer-to-peer platforms encourage a much higher rate of review than commercial sites		There is less self-selection bias in peer-to-peer ratings. Satisfaction is not the cause of positive bias in peer-to-peer ratings

Can selection bias explain the positive ratings in peer-to-peer platforms? Because online reviews are the main source of reputation for peer-to-peer services, platforms encourage a much higher review rate than commercial businesses. For example, Fradkin (2017) estimates that the review rate for trips on Airbnb is 70%, compared to 2.5% for Expedia. If consumers who experienced moderate satisfaction are least likely to provide a review, then the significantly higher review rate for peer-to-peer services means that more consumers with moderate opinions are included in the peer-to-peer ratings distribution. This should result in *less* extreme ratings in peer-to-peer services, not *more* extreme. Thus satisfaction is not likely to be the main driver of the positive bias in peer-to-peer

ratings. In sum, and as described in P_1, I anticipate that a consumer's evaluation of their service experience through expectancy disconfirmation, and their resultant satisfaction, does not affect peer-to-peer ratings in the same way as it does for commercial services.

> **P_1:** Consumers rate their providers based on an evaluation of perceived quality compared to initial expectations. For commercial services, the satisfaction that results from this evaluation is the primary driver of ratings. In contrast, for peer-to-peer services, expectancy disconfirmation evaluations affect ratings outside of satisfaction.

2.3   Trust and its Impact on Peer-to-Peer Ratings

I argue that expectancy disconfirmation may lead to trust, in addition to satisfaction. I further argue that contextual differences between peer-to-peer services and commercial services causes trust to have a stronger impact on peer-to-ratings than satisfaction, and causes consumers to rate peer-to-peer services positively even if expectations are merely met (vs. exceeded) or sometimes if they are negatively disconfirmed. See Table 2 for a summary of these contextual differences.

Table 2: Contextual Differences Between Peer-to-Peer and Commercial Services

Differences between Peer-to-Peer and Commercial Services	Reason for Differences
Peer-to-peer services have more uncertainty in performance outcomes	Peer providers are unknown to consumers. Heterogeneity in peer provider quality, and the lack of implicit quality signals make performance more difficult to predict.
Peer-to-peer services have more perceived risk	Peer providers are not professionals and can join platforms with limited vetting. Providers and consumers share private space together with limited oversight by the platform.
Ratings are considered to be more important to peer providers	Platforms match peer providers with potential customers in the membership network. Ratings are the primary driver of this matching. Platforms can encourage or prevent matches based on whether providers meet ratings thresholds.
Peer-to-peer consumers feel a higher need to justify their ratings	Peer providers pay close attention to ratings because of their importance. Consumers know that a low rating may make it difficult for the provider to acquire new business.
Standards of evaluation are relatively unclear in peer-to-peer services	Peer provider offerings are more heterogenous and not easily compared.
Causality for missed expectations is more difficult to assess in peer-to-peer services	Standards of evaluation are relatively less clear. Consumers are unsure whether a non-professional provider should be held to the same standard as a commercial provider.
Controllability for missed expectations is more difficult to assess in peer-to-peer services	Peer providers are not professionals. They may not have the skills, experience, or resources to control performance issues in the same way as a commercial provider.
Peer-to-peer consumers feel more gratitude and empathy toward their providers	Peer-to-peer sharing is a blend of economic and social exchange. Providers and consumers who share space and personal items together may develop social connections.

As previously discussed, consumers compare perceived quality against their initial

expectations. Spreng et al. (1996) demonstrated that this evaluation contains two

comparisons: a comparison of perceived quality with *desired* quality ("desire congruency"), and a comparison of perceived quality with *expected* quality ("expectation congruency"). Expectation congruency indicates how close the match was between what the consumer expected and what the product delivered, regardless of how well it met the consumer's desires. In other words, one may expect a product to perform relatively poorly, and if it did, it would be a match for expectation congruency (but not a match for desire congruency). Spreng et al. (1996) argue that expectation congruency leads to "information satisfaction"; a measure of how satisfied the consumer is with the claims given by the provider on which the consumer based their expectations. I argue that it can also lead to trust.

The link between expectancy disconfirmation and trust has not been extensively theorized in extant research, although the relationship is somewhat implied in the various definitions of trust. I propose that trust in exchange relationships is directly related to whether a provider honours their promises and commitments. Delivering on commitments is so important that many researchers have incorporated this concept into the definition of trust. For example, Sirdeshmukh et al. (2002, p. 17) define trust as "expectations held by the consumer that the service provider is dependable and can be relied on to deliver on its promises." Agustin and Singh (2005, p. 97) define trust as "a consumer's confident beliefs that he or she can rely on the seller to deliver promised service." Embedded in these and other definitions of trust (e.g., Lewicki et al., 2006; Mayer et al., 1995; Robinson, 1996; Rousseau et al., 1998) is the concept of expectations, and a belief or confidence that these expectations will be met (see Table 3 for a summary

of trust definitions and further theorizing that implicitly supports my argument for the link between expectancy disconfirmation and trust). By embedding the concept of expectations into the definition of trust, it implies that trust is built over time when expectations are continually confirmed (Zhang et al., 2020). Confirmed expectations lead to a confidence that exchange partners will deliver upon future commitments and expected positive outcomes. However, when promises are broken, and expectations are negatively disconfirmed, trust is lost (Lewicki et al., 2006; Lewicki & Bunker, 1995).

In the context of the dissertation, I use Morgan and Hunt's definition of trust as "confidence in an exchange partner's reliability and integrity" (1994, p. 23). Reliability suggests that past performance is a dependable predictor of future actions, while integrity means that exchange partners honor their commitments fairly and honestly (P. H. Kim et al., 2004). I argue that both reliability and integrity are related to meeting expectations. In the case of peer-to-peer exchange, consumer expectations for a peer provider are influenced by consumer reviews, and by information that the provider gives in their online profile and in direct communication with the consumer. When expectations are met (i.e., "confirmed expectations" in the vernacular of the expectancy disconfirmation model), it demonstrates that the provider is trustworthy. Specifically, reliability is demonstrated if prior performance (as documented in the consumer reviews) has been repeated, thereby confirming expectations. Integrity is demonstrated if a provider keeps the promises they made in their online profile and in direct communication, again confirming expectations.

Table 3: Selected Definitions of Trust and Possible Link to Expectancy Disconfirmation

Author	Trust Definition	Relevance to Expectancy Disconfirmation
Morgan and Hunt (1994)	"confidence in an exchange partner's reliability and integrity"	"The essence of opportunistic behavior is deceit-oriented **violation of implicit or explicit promises** about one's appropriate or required role behavior ... we posit that when a party believes that a partner engages in opportunistic behavior, **such perceptions will lead to decreased trust**".
Mayer, Davis and Schoorman (1995)	"willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor"	"When a trustor takes a risk in a trustee that **leads to a positive outcome, the trustor's perceptions of the trustee are enhanced.** Likewise, perceptions of the trustee will decline when trust leads to unfavorable conclusions."
Robinson (1996)	"expectations, assumptions or beliefs about the likelihood that another's future actions will be beneficial, favorable or at least not detrimental"	"Trust comes, in part, from judgments about integrity that are based on the perceived consistency of another's actions and the extent to which another's actions are congruent with his or her own ... **When promises are broken, this trust is shattered**".
Rousseau, Sitkin, Burt and Camerer (1998)	"a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another"	"**Reliability and dependability in previous interactions** with the trustor give rise to positive expectations about the trustee's intentions."
Sirdeshmukh, Singh and Sabol (2002)	"expectations held by the consumer that the service provider is dependable and can be relied on to deliver on its promises"	"**Benevolent behaviors provide diagnostic evidence of trust** because by going beyond the terms of the explicit 'contract', the service provider indicates pro-consumer motivations, restraint on self-serving opportunism, and a willingness to assume fiduciary responsibility"
Lewicki, Tomlinson and Gillespie (2006)	"[Psychological Trust] is confident expectations and/or willingness to be vulnerable"	"Trust grows with increased evidence of trustee's qualities, relationship history, communication processes, and relationship type and structural factors. **Trust declines when positive expectations are disconfirmed.**"

Although the link between expectancy disconfirmation and trust may be implied in many definitions of trust, the authors do not test this argument. However, Darke, Ashworth and Main (2010) test the relationship between expectancy disconfirmation and trust in the context of misleading advertising claims. They find that when consumers experience negative disconfirmation (operationalized as product performance that is worse than the advertised claims), consumers experience distrust. They also find that negative disconfirmation has a relatively stronger effect on trust than positive disconfirmation (i.e., when product performance is better than advertised claims). These results suggest that there is an opportunity to further establish the relationship between expectancy disconfirmation and trust in consumer research. Specifically, I will explore which factors strengthen the relationship between expectancy disconfirmation and trust, and subsequently, which factors affect the relationship between trust and word-of-mouth.

I propose that uncertainty and perceived risk are two important moderators in these relationships. Both factors are necessary for the development of trust (Bhattacharya et al., 1998), and both may be higher in peer-to-peer services than comparable commercial services. *Uncertainty* is the degree to which it is difficult to predict what will happen in a consumption experience, which often leads to a feeling of discomfort (Sun et al., 2012). Uncertainty is caused by a lack of knowledge about potential outcomes or by perceived variance in outcomes (Sitkin & Pablo, 1992). In the sharing economy, consumers transact with providers with whom they have no personal history. The providers may not have the skill to satisfy the consumer's needs, and their motives may be unclear (Belk, 2013). This contributes to a consumer's lack of knowledge. Further, the heterogeneity in peer-to-peer

offerings and quality contributes to perceived outcome variance.

When uncertainty is high, it provides the opportunity for exchange partners to demonstrate their trustworthiness (Molm et al., 2000). If outcomes are certain, then actions can be taken without the need for trust (Lewis & Weigert, 1985). Outcomes are rarely certain in practice, but even relative certainty can limit the impact of expectancy disconfirmation on trust. For example, imagine that a consumer decides to purchase a piece of meat from a butcher. The butcher has several locations around the city and has a very good reputation through consumer and expert reviews. The consumer has visited the butcher periodically in the past, and the product has consistently been good. If the consumer receives a good piece of meat on their next trip, and their positive expectations are confirmed, it should not drastically affect their perceptions of trust. The consumer already had trust in this butcher. If the consumer receives a bad piece of meat, they may be dissatisfied, but they likely won't question the butcher's trustworthiness. The butcher has demonstrated consistent quality in the past, and trust has been built over time.

On the other hand, imagine that a new butcher has moved into the neighbourhood. The consumer would like to try the new butcher, but they are uncertain if the butcher is better than their usual provider. The new butcher advertises "Top quality meat at fair prices". The consumer wonders if the quality will match the advertising. If the consumer receives a good piece of meat, it provides evidence that the butcher and their advertising can be trusted. If the consumer receives a bad piece of meat, they may question the provider's reliability and integrity. As demonstrated in this example, when outcomes are uncertain,

consumers search for information before committing to a transaction (Grant & Tybout, 2008). Often, consumers must rely on the service provider for this information (Singh & Sirdeshmukh, 2000). When relying on provider claims, consumers are attuned to evaluating whether performance meets commitments. If commitments are met, and the consumer's expectations are confirmed, the provider's trustworthiness is demonstrated. Thus:

> **P₂**: If a consumer is relatively uncertain that a provider is able and willing to deliver positive outcomes as expected, the provider's fulfillment of those expectations has a stronger impact on the consumer's perceptions of trust than if they were relatively certain about outcomes.

Peer-to-peer services also carry higher *perceived risk*. Providers are numerous and unknown to consumers, and can join a sharing platform with limited vetting. Thus the risk of opportunism may be higher than for commercial services. Initial meetings between consumers and peer providers take place online where trust is difficult to establish (Resnick & Zeckhauser, 2002). When exchanges move offline, they often lead to close personal interactions in private homes or vehicles, such as when an Airbnb consumer shares a home with the owner. The platforms have little control and oversight over these transactions as they are ongoing. This contributes to performance risk, safety risk, and property risk (Gullstrand Edbring et al., 2016; Schor, 2014).

Trust is related to risk because trust requires a willingness by exchange partners to make themselves vulnerable (Mayer et al., 1995). Vulnerability implies that if exchange partners act opportunistically, the consequences will be relatively severe (Chopra & Wallace, 2003). When risk is high, and the consequences of service failure are severe,

provider trustworthiness becomes more important. Consumers are primarily concerned that providers deliver on expectations, and that there are no surprises. For example, bungee jumping is a service experience with a very high risk, and consumers would want to ensure that their provider is completely trustworthy. Similarly, for peer-to-peer services, which carry a higher risk, it is important for consumers to evaluate the trustworthiness of their provider. In the absence of prior history with a provider, reputation serves as an antecedent to trust (Smith & Barclay, 1997) because it can signal that providers have performed without opportunism in previous exchanges (Cheema, 2008). For peer-to-peer services, ratings and reviews act as digital reputation. A consumer's evaluation of their trust in the provider should carry forward to the rating, as a signal to other consumers that this is a provider that can be trusted to honour their commitments. Thus:

P₃: If an exchange carries a relatively high perceived risk, the provider's trustworthiness is an important performance attribute which should be reflected in their rating.

P₄: Peer-to-peer services ratings will reflect the consumer's trust in the peer provider, in addition to the consumer's satisfaction with the experience.

The proposed link from trust to word-of-mouth is novel. Much more established is the link from word-of-mouth to trust. Electronic word-of-mouth (including online ratings and reviews) helps consumers to feel that a seller can be trusted. Indeed, the aggregate numerical rating, ratings valence, positive and negative framing of reviews, and quantity of reviews have been shown to affect a consumer's trust in a product or provider, and

lead to purchase intention, willingness-to-pay, and loyalty (Awad & Ragowsky, 2008; Ba & Pavlou, 2002; Casado-Aranda et al., 2019; Sparks & Browning, 2011; Utz et al., 2012). However, there is little research demonstrating a direct relationship from trust to word-of-mouth. My literature review finds three studies that show a positive relationship from trust to *self-reported* word-of-mouth. The contexts for the studies were telephone services (Ranaweera & Prabhu, 2003), banking and dentistry (Gremler et al., 2001), and supermarkets in Greece (Vlachos et al., 2011). These studies do not identify moderators for this relationship (such as perceived risk), nor do they measure the effect of trust on actual word-of-mouth behavior, which is a novel aspect of my dissertation.

Importantly, I believe that establishing trust as an outcome of expectancy disconfirmation, and establishing how trust affects ratings, should not be limited to peer-to-peer contexts. However, I believe that the peer-to-peer context is the right place to begin exploring the role and impact of trust because of the high risk and uncertainty (as previously discussed) and because of contextual factors that strengthen the influence of trust and lead to biased ratings, which I will explain next.

2.4   Confirmed Expectations and Positive Ratings Bias

To determine how evaluations of trust contribute to the positive ratings bias in peer-to-peer services, we return to our anecdote about Dave's experience with Airbnb:

Dave decided to use Airbnb. He booked a room with Mary after reading the consumer

reviews and looking through the pictures in the online listing. The experience was pretty much as he expected. There were no surprises; the accommodation was as it was advertised in the listing. It was a smallish room in an older house that was in a nice quiet neighbourhood. After the stay, Dave receives an email from Airbnb asking him to rate the experience. He is unsure what to do. The room was comfortable, but it wasn't particularly inspiring. He had hoped that maybe it would have been a bit nicer, but it did its job. Mary seemed nice. She was a good host and Dave doesn't want to give a lower rating that would make it harder for her to find new guests.

When a consumer evaluates performance against expectations, there are three possible outcomes. First, expectations may be confirmed, when performance merely meets expectations or falls within a "zone of indifference" where performance is close enough to expectations to not trigger disconfirmation (Woodruff et al., 1983). However, if expectations are disconfirmed, it is because performance was either better (positive disconfirmation) or worse than expected (negative disconfirmation). It is informative to review each of these situations and to compare their effect on ratings in commercial and peer-to-peer services. This allows us to begin to understand why peer-to-peer ratings have a positive bias (see Figure 2 for a flow chart summary of expected results).
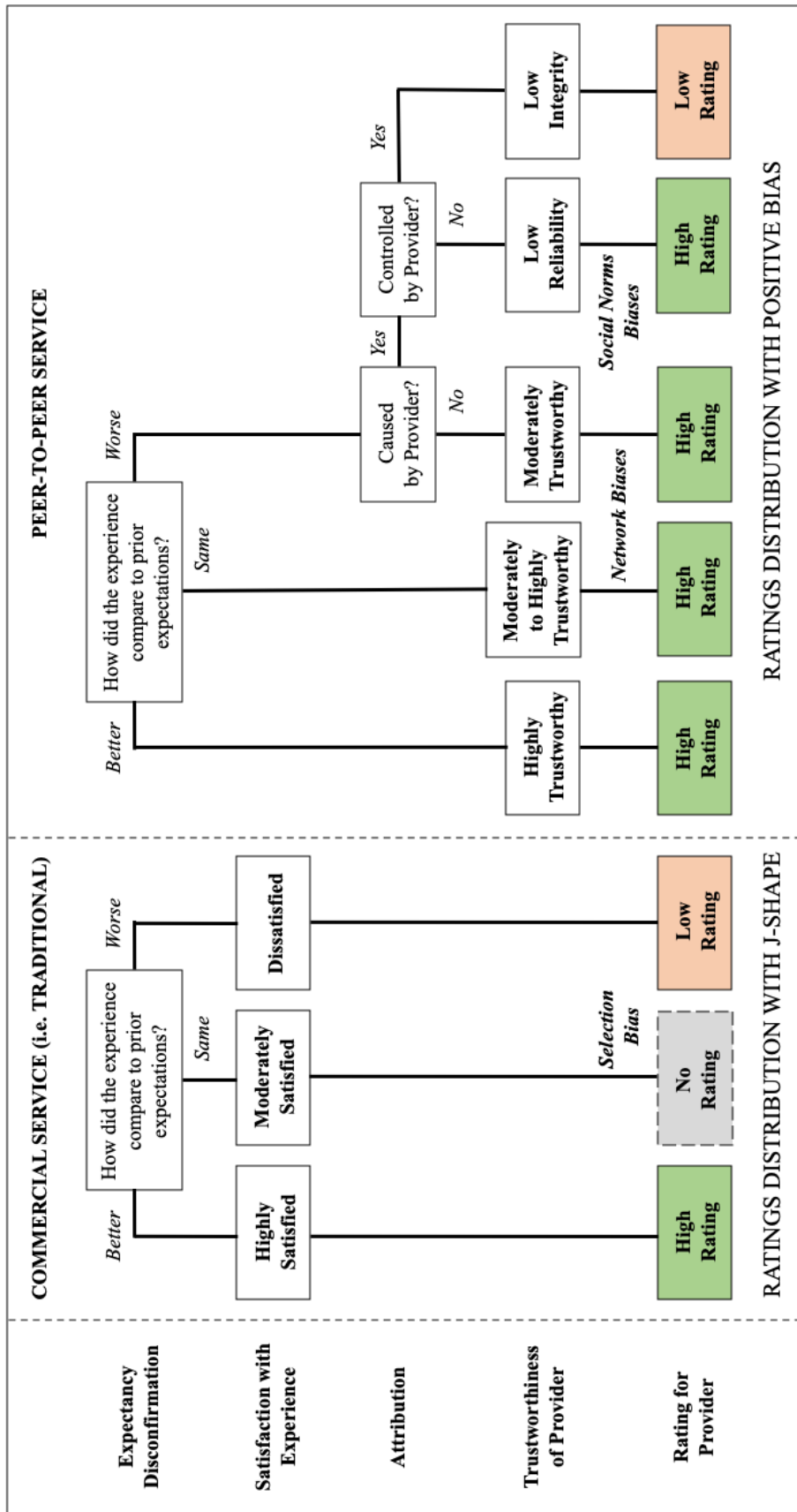
Figure 2: Relationship between Expectancy Disconfirmation and Ratings for Commercial and Peer-to-Peer Services

First, for commercial services, if performance exceeds expectations, the expectancy disconfirmation model suggests that it should result in a highly satisfactory experience with feelings of delight (Homburg et al., 2005). These high arousal emotions increase in intensity the further that performance exceeds expectation (Eisenbeiss et al., 2014), and lead to a desire to reward the provider with word-of-mouth (Anderson, 1998). The result is highly positive ratings. Similarly, for peer-to-peer services, if a provider exceeds expectations it should result in highly positive ratings. A provider that goes beyond their promised or contractual obligations signals that they care about the consumer (Sirdeshmukh et al., 2002). The provider's benevolence leads to high levels of trust (Blois, 1999), and consumers should feel a desire to reward that trust.

On the other hand, when expectations are merely met (i.e., confirmed), which occurs in the majority of cases for commercial providers (Woodruff et al., 1983), it does not trigger strong feelings of satisfaction and does not ignite a desire for word-of-mouth (Anderson, 1998). Self-selection leads to an underreporting bias; only consumers with extreme evaluations make the effort to post an online review (Schoenmüller et al., 2018). If commercial businesses could fix the underreporting bias, then the majority of ratings would be moderate, matching the prevalence of expectancy confirmation. This was experimentally demonstrated by Hu et al. (2017) who found that when commercial ratings were captured by all consumers, approximately 90% of the ratings were moderate (2, 3, or 4 stars). Thus moderate satisfaction leads to moderate ratings.

In contrast, when expectations are merely met in peer-to-peer services, it may lead to positive ratings (4 or 5 stars) through trust. There are two main reasons. First, even if expectations are merely met (rather than exceeded), trust may be relatively high. To demonstrate trustworthiness, consumers hold providers accountable for delivering on their promises; providers are not required to go above and beyond to meet a consumer's individual needs. Thus, when expectations are merely met, ratings that are related to trust should be higher than ratings that are related to satisfaction, because satisfaction is moderate when expectations are merely met.

P5: When expectations are confirmed, trust will be higher than satisfaction.

Second, network effects that are unique to peer-to-peer services decrease the importance of satisfaction (relative to trust) on ratings, and further compel consumers to increase their ratings. Peer-to-peer transactions are usually contained within a platform's network. Peer providers have limited resources to promote their business outside of this network (Benoit et al., 2017), and the reputation system is their primary means of differentiation and customer procurement. Platforms such as Airbnb use ratings to promote better providers, and can punish providers who fall below a ratings threshold, which jeopardizes their business (Jenkins, 2018). Thus, consumers should feel that ratings are more important to peer providers compared to commercial providers. Consumers may feel pressure not to harm a provider for a moderately satisfying experience, as long as the provider put forth a reasonable effort (i.e., as long as they were trustworthy). This leads to higher ratings when ratings are considered to more important to providers.

Most peer-to-peer platforms also have a two-sided rating system in which both the consumer and provider are rated. Both the consumer and provider must agree to an exchange, and the ratings and reviews are the primary basis on which exchange decisions are made. Because ratings are so important, peer providers monitor them closely. Consumers may be afraid that if they give a low rating, the provider will react with negative public feedback on the platform or negative private feedback to other network providers. This is worrisome, because consumers want to be seen as a fair exchange partner so that they will be able to attract future providers (Mulshine, 2015).

The relative importance of ratings and the fear of retribution from providers causes peer-to-peer consumers to feel that they may need to justify their ratings decisions. In turn, this causes the effect of satisfaction on ratings to be reduced. Satisfaction is concerned with needs fulfillment, which is somewhat subjective to an individual's unique needs and tastes, and so may be more difficult to justify. Satisfaction is also an emotional response (Oliver, 2010), rather than being based solely on cognitive judgements and hard facts. When consumers feel a need to justify, they tend to rely on reason-based rather than feeling-based decision making (Hong & Chang, 2015). One of the easier reasons to evaluate is whether the provider met their promised commitments and is trustworthy. If an Airbnb provider met their commitments, but the consumer was dissatisfied because the accommodation simply wasn't luxurious enough for their individual taste, it would be difficult to justify a low rating. The provider fulfilled their promises and therefore "did nothing wrong". In combination, the network effects of ratings importance and need to justify may contribute to a positive bias as described in the following propositions:

P$_{6a}$: Because satisfaction is relatively subjective, the effect of satisfaction on ratings will be reduced when consumers feel that ratings are very important to peer providers and when they have a strong need to justify their ratings.

P$_{6b}$: When ratings are considered to be very important to their providers, consumers will post higher ratings for their provider than they post when ratings are considered less important.

2.5   Negative Disconfirmation and Positive Ratings Bias

2.5.1   The Effect of Provider Causality on Trust

Figure 1, Panel B, illustrates the effect of negative disconfirmation on ratings. Negative disconfirmation occurs when performance is worse than expected. This should result in low satisfaction, trust, and ratings. However, as we will see, negative disconfirmation may result in high peer-to-peer ratings in some cases. Returning to our anecdote, let's imagine that Dave's experience with Airbnb did not go as well as he had expected:

Dave decided to use Airbnb. He booked a room with Mary after reading the consumer reviews and looking through the pictures in the online listing. However, the experience was not as good as he expected. The room was smaller and the home was older than it looked in the pictures. There was paint peeling from some of the walls, and the furniture was in need of an update. There was also no coffeemaker, which was usually a deal breaker for Dave. He was sure that he'd read everything in the listing, but maybe he'd

missed that. After the stay, Dave receives an email from Airbnb asking him to rate the

experience. He is unsure what to do. The experience was not great, but Mary seemed

nice. The home was clean and Mary did everything she could. He would feel bad rating

her poorly for things (like the room size) that were not under her control.

If a commercial service fails to meet a consumer's expectations, it causes dissatisfaction (Oliver, 2010). Compared to positive disconfirmation, "negative experiences are thought to be more salient, are perceived with greater intensity, and are expressed with a greater variety" (Oliver, 1993, p. 422). Dissatisfied consumers engage in more word-of-mouth than satisfied consumers (Anderson, 1998), and give more extreme evaluations (Eisenbeiss et al., 2014). This is because expectations act as a reference point against which consumers can evaluate performance (Homburg et al., 2005). Performance above the reference point is evaluated as a gain, while performance below expectations is a loss. Because losses loom larger than gains (Kahneman & Tversky, 1979), negative disconfirmation leads to stronger dissatisfaction (Mittal et al., 1998).

The fact that expectations are a reference point also explains why consumers do not feel a strong satisfaction response when performance merely meets expectations. However, when expectations are negatively disconfirmed, it triggers a search for attribution (Oliver, 2010). Consumers want to know what caused the missed expectations. Attribution theory explains how attributions are made (Heider, 1958); individuals evaluate causality and controllability before assigning blame (Weiner, 1985). *Locus of causality* refers to whether the cause is believed to be internal or external. In a consumption context,

consumers can attribute missed expectations internally to their own error in expectation-setting or product selection, or externally to the provider (Folkes, 1988). *Controllability* refers to volition. Was the performance failure controllable (and therefore preventable) or was it out of the control of the provider? Consumers engage in negative word-of-mouth when poor performance is deemed to be caused by and controlled by the provider. This generates anger and a desire to punish (Curren & Folkes, 1987).

I propose that locus of causality and controllability moderate the relationship between negative disconfirmation and the two dimensions of trust: reliability and integrity. When consumers experience missed expectations, they first seek to determine causality. If the cause is determined to reside with the provider (as it often does), then the provider is unreliable because they did not demonstrate an ability to reliably deliver performance at a level that meets the consumer's needs. However, causality may be more difficult to assess for peer-to-peer services. To determine causality of a missed expectation, consumers may consider whether their expectations were fair. Fairness of expectations can be evaluated by comparing them against an industry standard, which establish the performance level that the provider *should* have been able to deliver (Woodruff et al., 1983). Consumers generally have a good understanding of commercial standards, but they may be less clear about what standard to expect from a peer provider (Tussyadiah, 2016). For example, the hotel industry has well-established standards for attributes such as cleanliness and privacy. Consumers can form reasonable expectations for a budget, mid-range or luxury hotel. If performance fails to meet these competitive standards, a consumer can easily recognize the failure, and would likely attribute the cause and blame to the provider.

However, for peer-to-peer rentals on platforms like Airbnb, the types of accommodations are more varied. It could be a shared room with another guest, a private room in a shared home with the owner, or a full-home rental. The home may be the provider's personal residence, or it may be a separate property used for rentals only. It may be located in an apartment building, or in a detached home, or in the basement of a home, or in a castle, or in a treehouse (see Appendix B for examples of Airbnb accommodations). There may be differences in which rooms can be accessed during the rental, and what appliances are on-site. How can consumers easily and fairly compare one accommodation against another? When a provider invites guests into their home to share space with them, what level of service should be expected? With hotels, consumers clearly expect a spotlessly clean room. But should an Airbnb provider be expected to deliver hotel-level cleanliness in a home that they live in with their family? How harshly should the provider be judged?

When standards are less clear, consumers have difficulty assessing causality. Was the cause of negative disconfirmation due to poor performance or to expectations that were set too high or otherwise inaccurate? If a consumer is unclear on the standard that should be expected from a peer provider, how can they make a definitive determination that the provider has failed? The consumer may fault themselves for not being able to clearly set expectations, failing to properly assess the online listing, failing to ask the correct questions, and so on. The provider may therefore not be deemed unreliable.

> **P7:** If a consumer is not certain that a provider caused a performance failure, then the effect of *negative* disconfirmation on perceptions of provider reliability is weakened.

2.5.2   The Effect of Provider Control on Trust

Attributions of causality affect perceptions of reliability, but even if a provider was the cause of negative disconfirmation, it does not necessarily mean that the provider lacks integrity. I propose that controllability is an important factor in the determination of integrity, and that controllability may be more difficult to assess for peer-to-peer services. If a peer provider's service is worse than expected, but not egregiously low, consumers may consider the issues uncontrollable. This is because peer providers are not professionals, and may not have the aptitude, knowledge, or financial resources to fix negative issues. For example, an Airbnb provider may rent out a home in a less-than-ideal location, or a home that shows the wear and tear of several years of family use. The consumer may recognize that the issues with the home are the cause of their negative experience, but feel that the issues were not controllable. An Airbnb provider cannot change the age or location of their family home in the same way that a hotel provider would be expected to renovate its rooms when they become outdated. If negative disconfirmation is caused by the provider, but not under their control, the provider is unreliable, but not without integrity. The provider put forth a good faith effort, but did not have the ability to deliver expected quality.

On the other hand, if a consumer infers that a provider had control over the negative outcome, they may question the provider's integrity. For example, if an Airbnb provider misrepresents their home in the online listing, or fails to alert the consumer about an important appliance that is no longer functioning, then the missed expectations were controllable. Controllability suggests intentionality, and attribution theory suggests that

the consumer should hold the provider responsible for their actions (Oliver, 2010; Weiner, 1995). The consumer may determine that the provider lacks integrity because they attempted to deceive, and did not demonstrate a desire to fairly honour their commitments.

> **P8:** If a consumer is not certain that a provider had control over a performance failure, then the effect of *negative* disconfirmation on perceptions of provider integrity is weakened.

2.5.3   The Role of Social Norms Biases on Trust and Ratings

Peer-to-peer sharing blends economic with social exchange (Sundararajan, 2019). Although most peer-to-peer services are paid, providers are usually not professionals, and they invite consumers into their homes or share access to personal property. Further, providers and consumers often share space together which fosters social relationships (Celata et al., 2017). This mix of social and economic exchange may contribute to the positive ratings bias in peer-to-peer services. For example, Fradkin and colleagues (2015) suggest that consumers may give higher ratings to providers through reciprocity brought on by feelings of social connectedness. However, this proposition was not directly tested, and there is an opportunity to further explore the mechanism of the potential effect.

I propose that the blend of social and economic exchange in peer-to-peer services leads consumers to feel gratitude and empathy toward their provider. In turn, these positive emotions generate a desire to forgive peer providers for unreliable service. Consumers may feel gratitude toward their peer provider for their hospitality and for being invited

into the provider's personal space (Ikkala & Lampinen, 2015; Stofberg et al., 2019).

Social norms generate a desire to repay this hospitality. In social exchanges, benefits are

given based on need, and without expectation of reciprocation (Clark & Mills, 2011).

However, reciprocations made out of gratitude are appreciated (Aggarwal, 2004). In the

case of peer services, gratitude can be expressed by giving a positive rating to forgive

unreliable providers for issues that were out of their control. Consumers of peer-to-peer

services may also feel empathy toward providers because they are not professionals and

they are relatively independent from the platform (Costello & Reczek, 2020). This causes

consumers to think about the consequences of their actions toward the provider rather

than to the platform. Recognizing the importance of ratings to providers, consumers may

not want to harm a peer provider with a low evaluation (Filippas et al., 2018).

In summary, I propose that gratitude and empathy weaken the effect on unreliability on

ratings. On the other hand, if consumers infer that a provider lacks integrity, the

moderating effects of gratitude and empathy should be diminished. This is because the

provider prioritized their own self-interests over attempting to meet the consumer's

needs. The social exchange relationship has been broken, and social norms are no longer

applicable. Individuals tend to weigh negative information about integrity more strongly

than negative information about reliability (Kim et al., 2004). Thus, rather than forgiving

the provider, consumers may instead have a desire to punish them with a negative online

rating as a signal that the provider cannot be trusted.

> **P$_{9a}$:** When gratitude and empathy are high, consumers will post higher ratings for
> their provider than when gratitude and empathy are lower.

P$_{9b}$: Gratitude and empathy will cause consumers to forgive a provider for unreliable service, but not if the provider is deemed to lack integrity. Thus, when *negative* disconfirmation occurs, gratitude and empathy weaken the effect of reliability on ratings but not the effect of integrity on ratings.

In summary, I propose that peer-to-peer services are evaluated based on a consumer's trust in the provider, in addition to their satisfaction with the experience. Trust assessments are based on whether a provider meets expectations, and are affected by uncertainty and risk. Network effects (ratings importance and need to justify ratings) and social norms (gratitude and empathy) may cause consumers to give positive ratings to peer providers even if performance merely meets expectations, or sometimes, even if performance is worse than expected. Indeed, I propose that consumers may give low ratings only when they believe that a provider lacks integrity by not attempting to conduct a fair transaction. In the next section, I describe the research studies that were designed to test these propositions (Studies 1-5). Finally, in Study 6, I examine a new system of evaluation that I think can improve on the problems documented in the other studies. Specifically, I propose and test a novel ratings system to attenuate the positivity bias by making it easier to determine whether a provider is fair, honest, and trustworthy.

Chapter 2

3   The Current Research

The positive bias in peer-to-peer ratings is an important problem for platforms, and one that has not been sufficiently explored or explained. I propose that the high risk and uncertainty in peer-to-peer exchanges leads consumers to evaluate and rate their experiences based on their trust in the peer provider in addition to their satisfaction with the provider's performance. Contextual factors in the peer-to-peer sharing economy change how trust and satisfaction affect ratings, such that ratings may be positive when a consumer's expectations are merely met (rather than exceeded) or even if performance is worse than expected. These propositions will be tested in Studies 1 through 6. However, I first conducted a Pretest Study to confirm my assumptions about the important contextual differences between peer-to-peer and commercial services, before testing how these differences affect trust, satisfaction and ratings in the main studies.

4   Pretest Study

The Pretest Study was designed to compare peer-to-peer and commercial services across a variety of contextual factors that could affect how service experiences are evaluated, and how evaluations affect the online rating. Specifically, the tested contextual differences are summarized in Tables 1 and 2.

4.1 Method

4.1.1 Participants & Design

Two hundred and four North American participants were recruited via the Prolific online research panel. Participants were asked if they had purchased peer-to-peer services in the categories of short-term accommodation rental (e.g., Airbnb) or ride services (e.g., Uber) in the past year and to answer questions about their most recent purchase. If participants did not have any peer-to-peer service experience in the past year, they were asked to recall and answer questions about their most recent service experience with a commercial business provider in one of those same categories (hotels and taxis). Twelve participants (5.9%) did not purchase either peer-to-peer or commercial services in either of the target categories in the past year, and were excluded from the study. The final sample included 192 participants (92 women; $M_{age} = 39.0$ years) of which 110 recalled a recent peer-to-peer service experience and 82 recalled a recent commercial service experience.

4.1.2 Measures

Participants responded to questions concerning their feelings about their most recent peer-to-peer or commercial service experience. All items were measured on seven-point scales unless otherwise noted. First, participants were asked to think about how they felt when they were booking or about to begin their experience, including their *expectations* ("How high were your expectations for the level of quality and service that you would receive;" 1 = Very Low Expectations, 7 = Very High Expectations) and *perceived risk*

("How risky did you feel that it was to use that provider?;" 1 = Not at all, 7 = Extremely).

Uncertainty was a two-item measure ("How certain were you about the level of quality

and service that you would receive?," "How confident were you that the provider would

deliver the quality and service that you expected;" 1 = Not at all, 7 = Completely;

reverse-coded; α = .74). Uncertainty and perceived risk are theoretically distinct, but

closely-related constructs. Uncertainty is concerned with the fact that outcomes may be

difficult to predict, and is based on perceived outcome variance or a lack of knowledge.

These uncertain outcomes may be positive or negative, but the feeling of uncertainty is

uncomfortable and something that consumers wish to avoid. Risk is concerned with the

perceived likelihood of a negative outcome, and the potential severity of that outcome.

Risk goes beyond performance risk, and could include safety risks, social risks, etc. To

confirm the distinctiveness of these two constructs, I performed a correlation analysis and

found that the two measures were only moderately correlated ($r = .33$, $p < .01$).

Next, participants responded to questions about their feelings toward the experience

itself. This included *satisfaction* ("How satisfied were you with your experience;"

1=Very Dissatisfied, 7=Very Satisfied) and e*xpectancy disconfirmation* ("How did the

experience compare to your prior expectations")*,* which was measured from "1 = much

worse than I expected" to "7 = much better than I expected" (the midpoint "about the

same as I expected" represented confirmed expectations). This operationalization follows

the most common form of the scale (Oliver, 2010) and is well-established in satisfaction

research (e.g. Patterson, Johnson, & Spreng, 1996; Tsiros, Mittal, & Ross Jr, 2004). The

measure was then converted into three separate binary variables: *Confirmed expectations*

(i.e., a rating of 4 on the 7-point Expectancy Disconfirmation scale; 0 = No, 1 = Yes), *positive disconfirmation* (i.e., a rating of 5, 6, or 7 on the 7-point scale; 0 = No, 1 = Yes), or *negative disconfirmation* (i.e., a rating of 1, 2, or 3; 0 = No, 1 = Yes).

Participants who experienced negative disconfirmation were asked a series of questions to indicate whether the missed expectations were more a result of their own unrealistic expectations, or whether the missed expectations were caused by the provider. Specifically, *provider causality* was a four-item measure (α = .73) that was rated on a seven-point semantic differential scale: ("My personal expectations were too high for this type of service/The provider's quality or service level was too low for this type of service," "My personal expectations for the service were not accurate/The provider's description of the service was not accurate," "I must have missed some important information in the service description/The service description must have been missing some important information," "The missed expectations were completely my fault/ The missed expectations were completely the provider's fault"). Participants were also asked to assess controllability for missed expectations. *Provider control* was a three-item measure (α = .94; "My poor experience was definitely controllable by my provider," "My poor experience was definitely preventable by my provider," "My poor experience was definitely avoidable by my provider;" 1= Strongly Disagree, 7 = Strongly Agree).

Participants next responded to measures designed to understand how consumers of peer-to-peer and commercial services make ratings choices. *Provided rating* was a binary measure of whether the participant provided an online rating (0 = No, 1 = Yes). *Rating*

was a measure of the actual rating that was given (from 1 star to 5 stars). This measure was then converted into a binary variable called *five-star rating* which took the value 1 if the rating was exactly 5-stars, and 0 otherwise. I created the five-star rating variable to determine the percentage of ratings that were five-stars, which provides some context for the extremity of the positive distribution. *Ratings importance* was a three-item measure ($\alpha = .80$) of how important are ratings to the provider ("Online ratings from individual consumers are extremely important to the service provider," "The provider monitors their online ratings extremely closely," "Online ratings from individual consumers will greatly affect the provider's future business;" 1 = Strongly Disagree, 7 = Strongly Agree). *Need to justify* was a five-item measure ($\alpha = .75$) of whether consumers felt a need to justify ratings choices to the provider ("If I leave a poor rating, then I might have to justify my ratings choice," "If I leave a poor rating, then I might have to explain my ratings choice," "If I leave a poor rating, then I might be criticized by the provider," "If I leave a poor rating, then I might be punished by the provider," "If I leave a poor rating, then other providers might not want to service me;" 1 = Strongly Disagree, 7 = Strongly Agree).

The next set of questions measured what makes a peer-to-peer or commercial service worthy of a five-star rating for the attributes of overall quality (*quality 5-star*; 1 = As long as the quality was adequate and there were no major issues during the experience then it deserves a 5-star rating, 7 = The quality and service must be exceptional in every way to deserve a 5-star rating), condition of the rental asset (*condition 5-star*; 1 = As long as the provider is honest in their description of the [home/vehicle]'s condition even if it is worn out or in relatively poor condition then it deserves a 5-star rating, 7 = The [home/vehicle]

must look and work like it is brand new in order to deserve a 5-star rating), cleanliness

(*cleanliness 5-star*; 1 = As long as the [home/vehicle] is relatively clean then it deserves a

5-star rating, 7 = The [home/vehicle] must be spotlessly clean to deserve a 5-star rating),

responsiveness of the provider (*responsive 5-star*; 1 = As long as the provider responds to

any major concerns in a reasonable amount of time and makes some effort to try to fix

them then they deserve a 5-star rating, 7 = The provider must respond immediately and

fix all of my concerns in order to deserve a 5-star rating), and friendliness of the provider

(*friendly 5-star*; 1 = As long as the provider isn't rude then they deserve a 5-star rating, 7

= The provider must be exceptionally friendly to deserve a 5-star rating).

Next, participants indicated whether they were clear on the standards of expectation

(clarity of standards; "How clear are you about what are the standards of quality and

service that you should expect from a [X] service?;" 1 = Not at All, 7 = Completely).

Finally, participants indicated their gratitude and empathy toward the provider (i.e., *social

norms*) which was a six-item measure (α = .90) adapted from established scales (S. Lee et

al., 2014; Morales, 2005). Participants were asked "When thinking about your service

experience, to what extent did you feel the following emotions?": grateful, appreciative,

sympathetic, warm, compassionate, close (1 = Not at All, 5 = A Great Deal).

4.2 Results

The Summary Statistics and Correlation Matrix for the Pretest Study are in Appendix C.

Descriptive statistics by condition are in Table 4. ANOVA and logistic regression were

used to compare how consumers of peer-to-peer and commercial services evaluate their experiences. This comparison was performed at the overall level, and also within the specific service categories. As expected, there was no difference in the overall level of expectations for consumers of peer-to-peer and commercial services ($M_{Peer} = 4.55$, $M_{Commercial} = 4.46$, $F(1, 190) = .25$, $p = .62$). Also as expected, perceived risk was higher for peer-to-peer services ($M_{Peer} = 2.92$, $M_{Commercial} = 2.37$, $F(1, 190) = 5.88$, $p < .05$). This difference was driven by ride services ($M_{Peer} = 2.87$, $M_{Commercial} = 2.22$, $F(1, 111) = 5.41$, $p < .05$). However, uncertainty did not significantly differ between peer-to-peer and commercial services ($M_{Peer} = 3.03$, $M_{Commercial} = 2.82$, $F(1, 190) = 1.47$, $p = .23$).

Satisfaction was higher for peer-to-peer services ($M_{Peer} = 5.94$, $M_{Commercial} = 5.51$, $F(1, 190) = 4.07$, $p < .05$), but it did not significantly differ between peer-to-peer and commercial services for either accommodation or ride services individually. Expectancy disconfirmation was higher for peer-to-peer ($M = 4.77$) compared to commercial services ($M = 4.26$, $F(1, 190) = 7.11$, $p < .05$). This difference was significant for ride services ($M_{Peer} = 4.74$, $M_{Commercial} = 4.22$, $F(1, 111) = 4.98$, $p < .05$). The proportion of consumers who had their expectations merely met did not differ between peer-to-peer (42%) and commercial services (48%; $\beta = -.23$, $SE = .29$ $p = .43$). However, consumers of peer-to-peer services were significantly more likely to experience positive disconfirmation (48% vs. 33%; $\beta = .64$, $SE = .30$, $p < .05$) and marginally less likely to experience negative disconfirmation (10% vs. 20%; $\beta = -.78$, $SE = .42$, $p < .10$). Unexpectedly, provider causality ($M_{Peer} = 4.73$, $M_{Commercial} = 4.63$, $F(1, 25) = .05$, $p = .83$) and provider control ($M_{Peer} = 4.76$, $M_{Commercial} = 5.00$, $F(1, 16) = .11$, $p = .74$) did not significantly differ.

Table 4: Summary of Results for Pretest

	Total			Accommodation			Ride Service		
	P2P	Commercial	p value	P2P	Commercial	p value	P2P	Commercial	p value
	N = 110	N = 82		N = 42	N = 37		N = 68	N = 45	
Expectations	4.55 (1.30)	4.46 (1.17)	.62	4.88 (1.11)	4.70 (1.35)	.52	4.35 (1.38)	4.27 (.96)	.72
Perceived Risk	2.92 (1.56)	2.37 (1.57)	.02	3.00 (1.67)	2.54 (1.79)	.24	2.87 (1.50)	2.22 (1.36)	.02
Uncertainty	3.03 (1.11)	2.82 (1.21)	.23	2.93 (.87)	2.73 (1.39)	.44	3.09 (1.24)	2.90 (1.04)	.40
Satisfaction	5.94 (1.29)	5.51 (1.63)	.05	5.93 (1.28)	5.51 (1.59)	.20	5.94 (1.30)	5.51 (1.67)	.13
Expectancy Disc.	4.77 (1.30)	4.26 (1.36)	.01	4.83 (1.45)	4.30 (1.58)	.12	4.74 (1.22)	4.22 (1.17)	.03
Positive Disc.	.48 (.50)	.33 (.47)	.04	.50 (.51)	.43 (.50)	.55	.47 (.50)	.24 (.43)	.02
Confirmed Exp.	.42 (.50)	.48 (.50)	.43	.36 (.48)	.32 (.47)	.76	.46 (.50)	.60 (.50)	.14
Negative Disc.	.10 (.30)	.20 (.40)	.07	.14 (.35)	.24 (.43)	.26	.07 (.26)	.16 (.37)	.17
Provider Causality	4.73 (1.09)	4.63 (1.30)	.83	4.85 (1.18)	4.96 (.30)	.81	4.63 (1.71)	4.36 (1.12)	.75
Provider Control	4.76 (1.42)	5.00 (1.49)	.74	4.33 (1.39)	5.05 (1.60)	.48	5.33 (1.53)	4.92 (1.50)	.73
Provided Rating	.71 (.46)	.17 (.38)	.00	.76 (.43)	.16 (.37)	.00	.68 (.47)	.18 (.39)	.00
Rating (1 to 5 stars)	4.60 (.75)	3.93 (1.14)	.01	4.41 (.91)	4.00 (1.26)	.35	4.74 (.68)	3.88 (1.25)	.01
Five-Stars	.73 (.45)	.36 (.50)	.01	.59 (.50)	.50 (.55)	.68	.83 (.38)	.25 (.46)	.00
Ratings Importance	5.93 (.83)	5.22 (1.16)	.00	6.30 (.61)	5.30 (1.20)	.00	5.71 (.88)	5.15 (1.14)	.00
Need to Justify	4.29 (1.27)	3.97 (1.16)	.08	4.93 (1.10)	4.06 (1.00)	.00	3.89 (1.21)	3.88 (1.29)	1.00
Quality 5-star	3.78 (1.84)	4.62 (1.92)	.00	4.14 (1.77)	5.03 (1.89)	.04	3.56 (1.86)	4.29 (1.90)	.05
Condition 5-star	3.76 (1.93)	4.55 (1.91)	.01	3.88 (1.95)	5.08 (1.83)	.01	3.69 (1.93)	4.11 (1.89)	.26
Cleanliness 5-star	4.72 (1.92)	5.04 (1.71)	.24	5.00 (1.95)	5.49 (1.50)	.22	4.54 (1.90)	4.67 (1.81)	.73
Responsive 5-star	4.38 (1.83)	4.95 (1.84)	.18	4.38 (1.83)	4.95 (1.84)	.18			
Friendly 5-star	4.04 (2.02)	4.16 (1.92)	.77				4.04 (2.02)	4.16 (1.92)	.77
Clarity of Standards	4.92 (1.45)	4.77 (1.56)	.49	5.29 (1.20)	5.38 (1.34)	.75	4.69 (1.56)	4.27 (1.56)	.16
Social Norms	3.02 (.96)	2.71 (.87)	.02	3.08 (1.09)	2.68 (.91)	.08	2.98 (.87)	2.74 (.86)	.14

As expected, consumers of peer-to-peer services were significantly more likely to provide an online rating after their service experience ($\beta = 2.47$, SE $= .36$, $p < .01$). Overall, 71% of peer-to-peer experiences were rated but only 17% of commercial experiences were rated. This difference was significant for both accommodation (76% vs. 16%) and ride services (68% vs. 18%). Also as expected, the average rating for peer-to-peer services (M $= 4.60$) was significantly higher than for commercial services (M $= 3.93$, $F(1, 90) = 7.12$, $p < .05$). This overall difference was driven by the ride services category ($M_{Peer} = 4.74$, $M_{Commercial} = 3.88$, $F(1, 52) = 8.34$, $p < .05$). Five-star ratings (which represent the extreme positive end of the ratings scale) were also significantly more likely for peer-to-peer services ($\beta = 1.59$, SE $= .61$, $p < .05$). Overall, 73% of peer-to-peer ratings were five-stars, compared to only 36% of commercial ratings. This difference was again driven by the ride services category with 83% five-star ratings for peer-to-peer and only 25% five-star ratings for commercial ride services.

As expected, when making a ratings decision, consumers of peer-to-peer (vs. commercial) services believe that the rating is more important to their provider (i.e., ratings importance; $M_{Peer} = 5.93$, $M_{Commercial} = 5.22$, $F(1, 190) = 24.76$, $p < .01$), and they feel marginally more strongly that they may need to justify the rating to the provider ($M_{Peer} = 4.29$, $M_{Commercial} = 3.97$, $F(1, 190) = 3.19$, $p < .10$). Ratings importance is higher for peer-to-peer in both the accommodation ($M_{Peer} = 6.30$, $M_{Commercial} = 5.30$, $F(1, 77) = 22.79$, $p < .01$) and ride services category ($M_{Peer} = 5.71$, $M_{Commercial} = 5.15$, $F(1, 111) = 8.57$, $p < .01$), while need to justify ratings is significantly higher only in the accommodation category ($M_{Peer} = 4.93$, $M_{Commercial} = 4.06$, $F(1, 77) = 13.34$, $p < .01$).

When evaluating whether their service experience deserves a five-star rating, consumers of commercial services hold providers to significantly higher standards for quality (M_{Peer} = 3.78, $M_{Commercial}$ = 4.62, $F(1, 190) = 9.40$, $p < .01$) and for the condition of the rental asset (i.e., home or vehicle; M_{Peer} = 3.76, $M_{Commercial}$ = 4.55, $F(1, 190) = 7.81$, $p < .05$). However, the five-star ratings standards for cleanliness (M_{Peer} = 4.72, $M_{Commercial}$ = 5.04, $F(1, 190) = 1.41$, $p = .24$), responsiveness (M_{Peer} = 4.38, $M_{Commercial}$ = 4.95, $F(1, 77) = 1.86$, $p = .18$) and friendliness (M_{Peer} = 4.04, $M_{Commercial}$ = 4.16, $F(1, 111) = .09$, $p = .77$) did not differ between peer-to-peer and commercial services. Next, I expected that the standards against which consumers evaluate services would be less clear for peer-to-peer, but there was no difference (i.e., clarity of standards; M_{Peer} = 4.92 $M_{Commercial}$ = 4.77, $F(1, 190) = .47$, $p = .49$). Finally, as expected, social norms of gratitude and empathy were higher in peer-to-peer (M = 3.02) compared to commercial services (M = 2.71, $F(1, 190) = 5.24$, $p < .05$).

4.3 Discussion

The results of the Pretest Study were generally supportive of my predictions. In particular, the higher ratings in peer-to-peer services, coupled with a higher review rate, suggests that peer-to-peer ratings may be biased toward five-star ratings. As discussed, providing online ratings and reviews is effortful, and consumers who experience moderate levels of satisfaction are less likely to provide ratings than those who experience extreme satisfaction or dissatisfaction (i.e., self-selection, Schoenmüller et al.,

2018). Thus a higher review rate means that more consumers who experienced moderate satisfaction provided a rating, which should reduce the ratings average. The fact that ratings are higher (and five-star ratings are more likely) for peer-to-peer services despite the significantly higher review rate suggests that, as expected, self-selection is not a major cause of the biased positive distribution of peer-to-peer ratings.

This begs the question: what other factors may be impacting peer-to-peer ratings? Expectancy disconfirmation was significantly higher for peer-to-peer services, which led to higher satisfaction in the overall dataset (although satisfaction was not significantly different in either accommodation or ride services individually). This was unexpected, especially because, as predicted, the level of expectations did not differ between commercial and peer-to-peer services. Thus, differences in expectancy disconfirmation were not driven by lower expectations. As expected, perceived risk was higher for peer-to-peer services. Uncertainty did not differ, but this may be because the measure was retrospective. Where possible, the rest of the studies measure expectations, perceived risk, and uncertainty *before* participants experience the service, which is a more accurate reflection of how consumers feel in the moment.

Provider causality and provider control did not significantly differ, but there were relatively few observations for these variables because expectancy disconfirmation was primarily positive. Studies 4, 5, and 6 will attempt to elicit negative disconfirmation, which will allow for a better analysis of these variables and their effects on trust for peer-to-peer and commercial services. The results of the Pretest Study highlight other

contextual factors that could lead to higher ratings in peer-to-peer services, including the consumer's feeling that ratings are more important to peer providers, the need to justify ratings to peer providers, and social norms that encourage gratitude and empathy toward peer providers. The remaining studies will test how these variables affect the relationship between expectancy disconfirmation and ratings. Finally, the results of the Pretest Study suggest that there may be some important differences between the accommodation and ridesharing category. Future studies that examine more than one peer-to-peer category should include category-level dummy variables to control for these differences.

5   Study 1

The Pretest Study supported the prediction that ratings are higher in peer-to-peer services than commercial services, and that this is not a result of lower expectations. The review rate for peer-to-peer services was also higher, which suggests that the positive ratings distribution is not a result of self-selection, but rather, that individual consumer ratings may be biased. To determine the cause of this bias, we must understand how peer-to-peer consumers evaluate their service experiences, and how this may be different from consumers of commercial services. The objective of Study 1 is to test P_1, which proposes that for peer-to-peer services, unlike for commercial services, satisfaction does not fully mediate the effect of expectancy disconfirmation on ratings (see Figure 3). Because of the risk and uncertainty in peer-to-peer services, it is important that consumers have their expectations met, with no surprises. Therefore, expectancy disconfirmation has an added value in peer-to-peer services, beyond its effect on satisfaction. This proposition will be tested with the following hypotheses:

H₁.₁: For commercial services, the relationship between expectancy disconfirmation and ratings is fully mediated by satisfaction

H₁.₂: For peer-to-peer services, expectancy disconfirmation has both a direct effect and an indirect effect through satisfaction on ratings
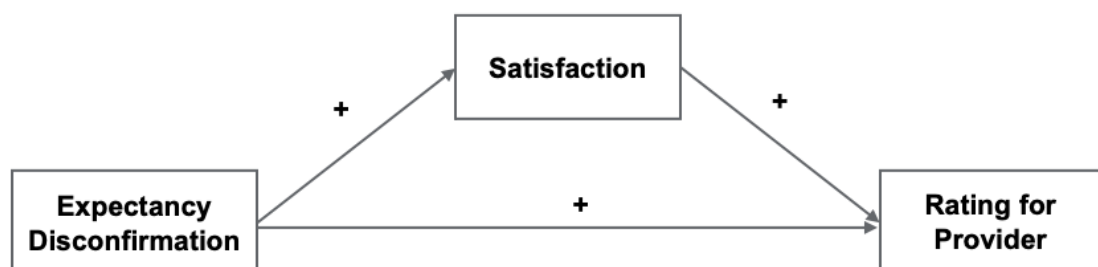
Figure 3: Model to be tested in Study 1

The context for this study is food service, which is a category that is familiar to the participants, and which could plausibly be delivered by either a commercial or peer provider. To simulate real-world ratings decisions, the study involved deception by having participants believe that they were rating an actual new food service business. It was important to have realism in the studies because the positive ratings bias is impacted by social and network factors (such as trust, gratitude and empathy, importance of ratings, and feeling that ratings need to be justified). These contextual factors cannot be easily replicated in a lab.

5.1   Method

5.1.1   Participants & Design

The experiment was conducted in the waiting area outside of a behavioral lab.

Participants (N=146) were undergraduate students who were at the lab to complete an

unrelated study for course credit. Participants were randomly assigned to one of two

conditions (*Service Type*: commercial vs. peer-to-peer). A control condition was later

added for additional information on the direction of results. While waiting to participate

in the unrelated study, participants were approached by a confederate with a tray of

chocolate chip cookies. The confederate was dressed in a blue t-shirt with a Home Café

logo (commercial business condition) or a plain blue t-shirt with no logo (peer-to-peer

and control conditions; see Appendix D). The confederate asked participants if they

would like to sample a chocolate chip cookie, and to take part in a survey about a new

food service called "Home Café" that had recently been approved by the university. We

used a fabricated company name and brand logo to control for the fact that existing

attitudes toward real brands could impact participant responses (Luffarelli et al., 2019).

To encourage variance in ratings, the cookies had been left out overnight to prevent their

freshness from creating a ceiling effect in participants' ratings of quality. The confederate

told participants that he baked the cookies himself that morning, either as an employee of

a new business (commercial condition), or as a provider for a new peer-to-peer service

(peer-to-peer condition). He subsequently explained that the new business sold and

delivered home-style meals and snacks to students on campus, or that the service allowed

students to sell and deliver their own homemade meals and snacks to other students on

campus in the manner of Airbnb or Uber. In the control condition, the confederate told

participants that the cookies were a new item for the campus cafeteria and that he was administering the survey on their behalf.

After sampling a cookie, participants were handed a paper survey. The survey first asked participants to provide their email address to receive discounts and free items from Home Café. Email addresses were captured to encourage participants to answer the survey thoughtfully, knowing that the confederate and the business could contact them about their ratings. This enhanced the realism of the rating experience. Participants rated the cookie, and answered questions about their prior expectations, satisfaction, and how the cookie compared to their expectations (i.e., expectancy disconfirmation). The confederate told participants to place their completed surveys in a closed box, and that he would take the surveys back to Home Café at the end of the day.

5.1.2   Measures

The dependent measure is the rating for the cookie on a 5-star ratings scale. Participants then retroactively rated their expectations (i.e., their level of expectations prior to sampling the cookie) on a 10-pt scale (1=very low, 10=very high). Because of the design of the experiment, and the need for deception, it was not possible to measure expectations prior to consumption. However, expectations were measured before satisfaction, following the direction from Oliver (2010). Satisfaction was measured on a 10-point scale ("How satisfied were you with your experience;" 1=very Unsatisfied, 10=very Satisfied). I used a one-item measure for satisfaction because I wanted the participants to believe that the paper survey and the questions had been created by the Home Café business

rather than for a research study. The one-item measure was used in the remaining studies

for consistency. Finally, expectancy disconfirmation was measured on a 5-pt scale from

"1 = much worse than I expected" to "5 = much better than I expected".

5.2   Results

5.2.1   Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 1 are in Appendix E.

Descriptive statistics by condition are in Table 5. As predicted, participants' expectations

did not significantly differ between the commercial and peer-to-peer conditions

($M_{Commercial}$ = 5.98, M_{Peer} = 6.17, $F(1, 124)$ = .36, p = .55). Expectations also did not

significantly differ between these conditions and the control condition (M = 6.50). There

was no significant difference between the commercial and peer-to-peer conditions for

rating ($M_{Commercial}$ = 4.33, M_{Peer} = 4.23, $F(1, 122)$ = .52, p = .47), satisfaction ($M_{Commercial}$

= 8.43, M_{Peer} = 8.57, $F(1, 123)$ = .30, p = .58), or expectancy disconfirmation ($M_{Commercial}$

= 4.11, M_{Peer} = 4.29, $F(1, 123)$ = .22, p = .14). The majority of participants in both the

commercial and peer-to-peer conditions (87%) indicated that they experienced positive

disconfirmation, which explains why ratings were similarly high in these two conditions.

Although the cookies were left out overnight to reduce freshness, the fact that they were

given out for free likely impacted the high levels of positive disconfirmation and

satisfaction. Taking this into account, later studies will be designed to elicit more

variance, and specifically more negative disconfirmation.

Table 5: Means and Standard Deviations for Study 1

	Commercial N = 66	Peer-to-Peer N = 58	Control N = 19
Expectations	5.98 (1.77)	6.17 (1.84)	6.50 (1.44)
Rating	4.33 (.72)	4.23 (.72)	4.11 (.59)
Satisfaction	8.43 (1.45)	8.57 (1.30)	8.34 (1.13)
Expectancy Disconfirmation	4.11 (.70)	4.29 (.67)	3.74 (.73)

5.2.2   Tests of Hypotheses

The expectancy disconfirmation paradigm predicts that product and service ratings should be affected by expectancy disconfirmation through satisfaction. To test this mediation, a series of bootstrap analyses were performed using 5000 samples and a 95% bias-corrected confidence interval (PROCESS Model 4, Hayes, 2018). Expectancy disconfirmation was the predictor, satisfaction was the mediator, and ratings was the dependent variable. As predicted, the indirect effect of expectancy disconfirmation on rating, through satisfaction, was significant for the control (β = .24, SE = .15, 95% CI: .01, .51), commercial (β = .39, SE = .11, 95% CI: .14, .62), and peer-to-peer conditions (β = . 39, SE = .12, 95% CI: .19, .66).

In this study, it was further predicted that expectancy disconfirmation should have an additional effect on rating in the peer-to-peer condition because of the importance of trust when dealing with an unknown peer provider. Supporting $H_{1.1}$, there was no direct effect of expectancy disconfirmation on rating for the commercial condition (β = -.07, SE = .09,

95% CI: -.25, .10). This replicates extant research using the expectancy disconfirmation model. There was also no direct effect in the control condition (β = .02, SE = .13, 95% CI: -.26, .30). However, supporting H1.2, there was a direct effect of expectancy disconfirmation on rating for the peer-to-peer condition (β = .28, SE = .11, 95% CI: .06, .51). The direct effect and indirect effect (through satisfaction) are in the same direction, demonstrating a complementary mediation (Zhao et al., 2010).

5.3   Discussion

The results of Study 1 support proposition P_1. For peer-to-peer services, expectancy disconfirmation has a direct effect on ratings beyond the mediated effect of satisfaction. The results show a distinction between peer-to-peer evaluations and commercial service evaluations, which follow the well-established expectancy disconfirmation process. The study further supports the assumptions outlined in Table 1; expectations are not significantly lower in peer-to-peer services, and ultimately, satisfaction is likely not the main driver of the positive bias in peer-to-peer ratings. As I have shown, in peer-to-peer services there is a direct route from expectancy disconfirmation to ratings, independent of satisfaction. A question still remains as to how this link operates. That is, how does expectancy disconfirmation affect ratings outside of its effect on satisfaction? Based on the theory and literature reviewed so far, I posit that trust is an important mediator in the peer-to-peer situation due to the higher uncertainty and perceived risk in peer-to-peer services. The effects of uncertainty and risk on trust will be tested in Study 2.

6 Study 2

Study 1 demonstrated that expectancy disconfirmation directly affects ratings for peer-to-peer services, in addition to the indirect effect through satisfaction. The objective of studies 2A and 2B is to demonstrate an important mechanism of this effect: *trust*. The studies will test propositions P2, P3, and P4. I propose that expectancy disconfirmation is more strongly related to trust when uncertainty is high (P2), because confirmed expectations demonstrate that the provider is able and willing to meet the consumer's needs. I expect that trust is more strongly reflected in ratings when perceived risk is high (P3), because when a service is risky, the provider's trustworthiness is a differentiating attribute that will be important for provider selection. Ratings in peer-to-peer services (which carry relatively higher uncertainty and risk than comparable commercial services) should thus be reflected by trust in addition to satisfaction (P4). I manipulate uncertainty in Study 2A to test P2. I then manipulate perceived risk in Study 2B to test P3 and P4. Table 6 lists the hypotheses to be tested and Figure 4 illustrates the model to be tested.

Table 6: Study 2 Hypotheses

2.1	Expectancy disconfirmation is positively related to trust in the service provider
2.2	Uncertainty moderates the relationship between expectancy disconfirmation and trust; this relationship will be stronger when uncertainty is higher.
2.3	Trust is positively related to ratings
2.4	Perceived risk moderates the relationship between trust and ratings; this relationship will be stronger when perceived risk is higher.
2.5	When uncertainty and perceived risk are high, expectancy disconfirmation is positively related to ratings, and is mediated by trust in the service provider.

6.1   Study 2A Method

Study 2A tested the first three hypotheses. I expected that expectancy disconfirmation is

related to trust, and that this relationship is stronger when uncertainty is higher, as it is in

peer-to-peer services. Further I expected that a consumer's trust would be reflected in

their ratings, per $H_{2.3}$ (see Figure 4).
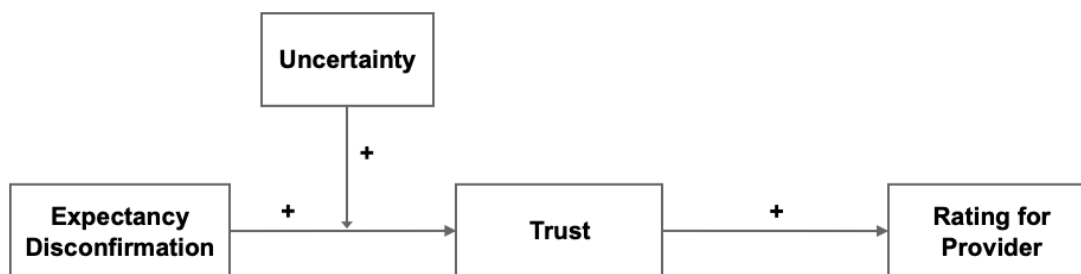
Figure 4: Model to be tested in Study 2A

6.1.1   Study 2A Participants, Design, and Measures

One hundred and eighty-seven undergraduate students were recruited for course credit

(80 women; M_{age} = 19.85 years). Participants were randomly assigned to one of four

conditions in a 2 (*Service Type*: commercial vs. peer-to-peer) x 2 (*Service Tenure:* new

vs. established) between-subjects design. The study was completed in two phases. In the

first phase, participants received an email from the lab manager with the instructions. The

email stated that the university was considering hiring a graphic design company

("Netwrk") to provide design services for students, and that the purpose of the study was to test the company's services for resume design. The company name and brand logo were fabricated to control for the fact that existing attitudes toward real brands could impact participant responses. In the *commercial conditions*, Netwrk was described as a service for graphic design (posters, business cards, résumés, logos etc.) In the *peer-to-peer conditions*, the description added that Netwrk was a peer-to-peer platform (like Airbnb and Uber) in which individual *freelance* designers can join the platform to be matched with potential customers. Further, in the *new conditions*, Netwrk was described as a *brand new* service that was currently in its pre-launch phase. The company had not yet launched to the public but had been working with prospective clients. Conversely, in the *established conditions*, Netwrk was described as being in business for the past two years and having an established client list. A provider's reputation (i.e., evidence of past performance) helps consumers to form expectations, and is especially important for services because it is difficult to judge expected quality until they are consumed. When providers are new they have no reputation; thus consumers in the *new conditions* should be relatively more uncertain about the quality that they will receive (Kim & Peterson, 2017; Lovett et al., 2013).

Participants were provided with a link to the first of two surveys, where they were asked about their feelings about the upcoming experience including their *expectations* (1 item), *perceived risk* (1 item), and *uncertainty* (two items; α = .70), which were measured the same as in the Pretest Study. Results of a correlation analysis showed that there was no correlation between perceived risk and uncertainty (r = .09, p = .25) which provides some

support that I was able to isolate uncertainty (and not risk) in the manipulation of Service Tenure. Finally, participants provided their email address and were told that a Netwrk designer would contact them via email later that day to begin the design process.

The second phase of the study began approximately one hour after the participant had completed the first survey. Posing as a Netwrk designer, I sent an email to the participant (addressed from michaelm.netwrk@gmail.com) to start the design process (see Appendix F). The email began with a boilerplate from Netwrk stating that the company had *assigned* a graphic designer ("Michael") to the project (commercial conditions) or that the company had *matched* the participant with a *freelance* designer to work on the project (peer-to-peer conditions). The boilerplate also included a short bio and a photograph of the designer. In the bio for the *new conditions*, the designer was described as having recently joined Netwrk. Further, Michael stated later in the email that this was one of his first design projects with the company. Conversely, in the *established conditions*, the designer was described in the bio as having completed over 100 projects with Netwrk. After the boilerplate, Michael introduced himself and explained that he needed some information from the participant: a copy of their current resume; some direction on the style of resume that they wanted including whether they preferred it to be classic or modern, and professional or artistic; and additional background information that could be used to customize the design. The answers to the style questions were used to select one of three resume design templates that would be used for the new design.

A research assistant transferred the information from the participant's original resume

into the selected template to create the new design. None of the information from the

participant's resume was changed. It was simply transferred into the new design template

(for an example, see Appendix G). We used three design templates for the new resume

rather than one because it was important for participants to feel that they were taking part

in a real design experience and receiving a customized design. However, the templates

may differentially affect the service evaluation so I created a dummy variable to identify

which template was given to each participant (0 = No, 1 = Yes). These were then used as

covariates for the hypothesis tests. The background information that the participants

provided in their email was not used to personalize the design. Rather, the act of giving

additional information was designed to increase the participant's engagement.

Participants responded to the email and provided the requested information. They

received a thank you reply from the designer and were told that the designer would

deliver their new design within 24 hrs, per Netwrk's service policy. The following day,

participants received a third email from the designer with the new resume. The email was

delivered approximately one hour after the end of the 24-hour window, thus failing to

meet the service commitment. The designer apologized, and stated that he was delayed

because he was working on another project. The service failure was designed to increase

the variance in expectancy disconfirmation based on the learning from Study 1 that

participants tended to evaluate the "free" service experiences favorably. The email

included a link to the second survey, on which participants could evaluate the service

experience. First, participants were asked to provide a *rating* on a 5-star scale. They were

told that the rating would be assigned to the designer, and would be shared with Netwrk

so that they could post it on their website. Participants next indicated *expectancy disconfirmation* (1 item) which was measured the same as in the Pretest Study. *Satisfaction* (1 item) was next measured on a seven-point semantic differential scale ("Very Dissatisfied/Very Satisfied"). Finally, participants were asked to evaluate their *trust* in the provider, which was a four-item measure ($\alpha = .90$) adapted from Sirdeshmukh et al. (2002). The scale included two items designed to measure provider reliability ("The designer is very reliable," "The designer is very competent;" 1 = Strongly Disagree, 7 = Strongly Agree) and two items designed to measure provider integrity ("The designer has very high integrity," "The designer can definitely be counted on to do what's right;" 1 = Strongly Disagree, 7 = Strongly Agree).

6.2   Study 2A Results

6.2.1   Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 2A are in Appendix H. Descriptive statistics by condition are in Table 7. Unexpectedly, the service tenure manipulation did not significantly increase uncertainty in the new (M = 2.90) compared to the established condition (M = 2.82 ; $p = .63$). I further expected that uncertainty would be higher for peer-peer services, but results showed that uncertainty did not significantly differ between the peer-to-peer (M = 2.90) and commercial conditions (M = 2.80 ; $p = .51$). The overall pattern of results were as expected; uncertainty was highest in the new peer-to-peer condition, and lowest in the established commercial service condition, suggesting that a stronger version of the same manipulation may be needed. Next, I

expected and found that perceived risk was higher in the peer-to-peer (M = 3.08)

compared to the commercial condition (M = 2.45; p < .01). As expected, perceived risk

did not significantly differ between the service tenure conditions (M_{New} = 2.62; $M_{Established}$

= 2.97; p = .12). Participants' expectations also did not significantly differ between any

of the four conditions.

Table 7: Means and Standard Deviations for Study 2A

	P2P New N = 48	P2P Established N = 53	Commercial New N = 46	Commercial Established N = 40
Expectations	5.26 (.94)	5.42 (.98)	5.50 (1.19)	5.40 (1.15)
Perceived Risk	2.98 (1.45)	3.17 (1.54)	2.23 (1.31)	2.70 (1.71)
Uncertainty	2.94 (1.09)	2.88 (.93)	2.85 (1.05)	2.75 (1.12)
Rating	4.02 (1.04)	4.17 (.83)	4.22 (.97)	3.95 (1.08)
Expectancy Disc.	4.22 (1.75)	4.44 (1.42)	4.58 (1.79)	4.13 (1.75)
Satisfaction	4.74 (1.61)	5.02 (1.15)	5.31 (1.38)	4.72 (1.65)
Trust	5.24 (1.29)	5.48 (.93)	5.75 (1.19)	5.34 (1.21)

6.2.2   Tests of Hypotheses

Next, hypotheses $H_{2.1}$ and $H_{2.2}$ were tested. Supporting $H_{2.1}$, there was a significant

positive relationship between expectancy disconfirmation and trust (β = .44, SE = .04, p <

.01). To test $H_{2.2}$, I performed two different analyses to demonstrate that the relationship

between expectancy disconfirmation and trust is stronger when uncertainty is higher.

Recall that I attempted to manipulate uncertainty through the service type and service

tenure conditions. Thus, in the first test of $H_{2.2}$, I expected to find a three-way interaction

such that the effect of expectancy disconfirmation on trust was highest for new peer-to-peer services, because of higher uncertainty in that condition. From a moderated moderation analysis (PROCESS model 3; Hayes 2018), there was a marginally significant 2-way interaction between expectancy disconfirmation and service tenure, such that the effect of expectancy disconfirmation on trust was stronger in the "new" condition ($\beta = .15$, SE $= .08$, $p < .10$). The 2-way interaction with service type (commercial vs. peer-to-peer) was not significant ($\beta = -.03$, SE $= .08$, $p = .71$). Further, providing support for $H_{2.2}$, there was a marginally significant 3-way interaction such the effect of expectancy disconfirmation on trust was highest in the new peer-to-peer condition ($\beta = .32$, SE $= .17$, $p < .10$). Next, a moderation analysis (PROCESS model 1; Hayes 2018) was conducted to *directly* test $H_{2.2}$ with expectancy disconfirmation as the predictor, trust as the dependent variable, and uncertainty as the moderator. There was a positive main effect of expectancy disconfirmation on trust ($\beta = .42$, SE $= .04$, $p < .01$) and a negative main effect of uncertainty on trust ($\beta = -.26$, SE $= .07$, $p < .01$). However, the interaction was not significant ($\beta = .06$, SE $= .04$, $p = .11$, d $= .25$).

Finally, I tested $H_{2.3}$. As expected from $H_{2.3}$, there was a significant positive relationship between trust and rating ($\beta = .56$, SE $= .04$, p $< .01$). To test this relationship further, I conducted a mediation analysis (PROCESS model 4; Hayes 2018) with expectancy disconfirmation as the predictor, trust as the mediator, and rating as the dependent variable. As expected, there was a significant indirect effect of disconfirmation on rating through the mediator trust ($\beta = .14$, SE $= .03$, 95% CI: $.08$, $.20$). This process will be further tested in Study 2B.

6.3   Study 2A Discussion

The results of Study 2A provide some support that expectancy disconfirmation is related

to trust, and that trust is related to ratings. However, the predicted interaction between

expectancy disconfirmation and uncertainty on trust (H$_{2.2}$) was not significant. There

were a few issues in this study that may have contributed to the non-significant result.

First, the manipulation for service tenure was not successful in creating significant

differences in uncertainty, and there also was no significant difference between the peer-

to-peer and commercial conditions. This may have been impacted by the fact that overall

levels of uncertainty across the conditions were relatively low. The fact that "Netwrk"

was altering the participant's existing resume rather than building a brand new resume

may have contributed to this result. Further, the effect size of the interaction between

expectancy disconfirmation and uncertainty on trust was relatively small. Some

participants may have guessed that they were participating in a scenario as part of a

research study rather than interacting with a real provider, which could have affected this

result. Although the deception in this study was well designed, it is not known how many

participants were deceived. Attempts were made to address these issues in Study 2B.

6.4   Study 2B Method

Study 2B was designed to test the effect of trust on rating (H$_{2.3}$) and the moderating effect

of perceived risk on this relationship (H$_{2.4}$). Further, the study tests the full moderated

mediation described in $H_{2.5}$. Specifically, I predicted that the effect of expectancy

disconfirmation on rating would be mediated by trust (in addition to satisfaction), and

that the mediation through trust would be strengthened when uncertainty and risk are

higher (see Figure 5).
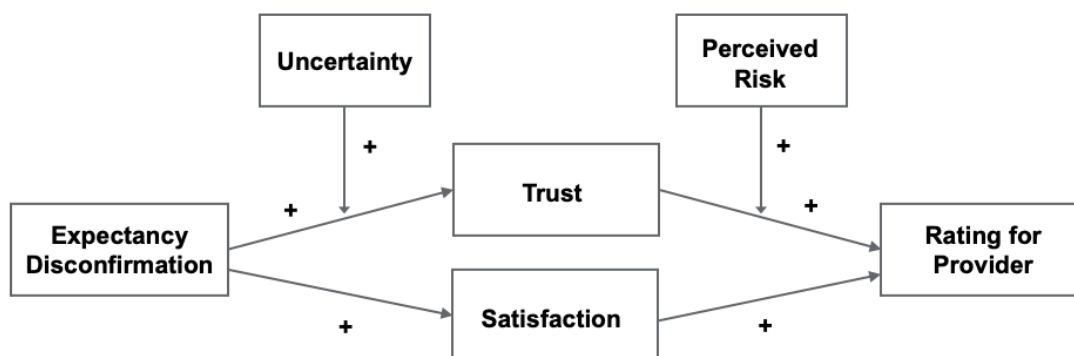
Figure 5: Model to be tested in Study 2B

6.4.1   Study 2B Participants, Design, and Measures

Two hundred and nine undergraduate students were recruited for course credit (105

women; M_{age} = 19.28 years). Participants were randomly assigned to one of four

conditions in a 2 (*Service Type*: commercial vs. peer-to-peer) x 2 (*Risk Level:* high vs.

low) between-subjects design. The study followed a similar design to Study 2A. The

same variables from Study 2A were measured in Study 2B in the same fashion.

Additionally, I included a measure to assess the effectiveness of the deception.

Participants were debriefed at the end of the study and were asked to indicate whether,

during the experience, they believed that they were interacting with a real design

company and designer ("When you were participating in this study, how much did you

believe that Netwrk was a real company and that a real designer was working on your

project?;" 1 = I felt completely sure that Netwrk was fake and this was only for research,

4 = I was not sure whether this was fake or real, 7 = I felt completely sure that Netwrk

and its designer were real). Nearly 90% of participants felt somewhat to completely sure

that the experience was real (i.e., a rating of 5, 6, or 7). The other 24 participants who

questioned the veracity of the experience were removed from the data to improve the

precision of the analyses (final sample of 185 participants, 91 women; M_{age} = 19.28

years).

As in Study 2A, the study was completed in two phases. In the first phase, participants

received an email from the lab manager with the instructions. They were told that they

would be testing a new service for logo design. Service type was manipulated in the same

fashion as in Study 2A. In the *commercial conditions*, the email from the lab manager

described Netwrk as a service for graphic design, while in the *peer-to-peer conditions*,

the description added that Netwrk was a peer-to-peer platform (like Airbnb and Uber) in

which individual freelance designers can join the platform. Further, in the *high risk*

conditions, the email from the lab manager included a prominent warning that Netwrk

was a third-party website that was not affiliated with the University. Participants were

told that by participating in the study, they would be sharing personal information with

this third-party website and a designer, and that the University could not guarantee the

privacy of the information. Participants were further told that by clicking on the link to

begin the first survey, they were acknowledging that they had read the warning and were accepting the risk. The same warning was included in the first survey. After reading the letter of information in the survey, participants read that by participating in the study, they would be sharing personal information with a third-party website, and that the University could not guarantee the privacy of the information. Participants clicked forward in the survey to indicate that they acknowledged the risk. These warnings were not included in the *low risk conditions*.

Participants completed the first survey to indicate their *expectations* (1 item), *perceived risk* (1 item), and *uncertainty* (two items; α = .75), and provided an email address through which they could be contacted by the designer. Results of a correlation analysis showed that there was a weak correlation between perceived risk and uncertainty (r = .18, p < .05) which provides some support that I was able to isolate risk (and not uncertainty) in the manipulation of Risk Level. The second phase of the study began approximately one hour after the participant had completed the first survey. Posing as a Netwrk designer, I sent an email to the participant including a boilerplate with a short bio and photograph of the designer. The boilerplate stated that Netwrk had *assigned* a graphic designer (commercial conditions) or stated that the participant had been *matched with* a freelance designer (peer-to-peer conditions). After the boilerplate, the designer introduced himself and explained that he needed some information from the participant to start the design process: their first and last name along with a key phrase or motto to include in the design; some direction on the style of logo that they wanted including whether they preferred it to be colourful or neutral, and clean/simple or detailed/artistic; and to

highlight their interests from a list of choices including sports, music, art and business, which would help the designer determine a theme for the logo. The answers to the style and interest questions were used to select one of three design templates that would be used for the new logo. Because the templates may differentially affect the service evaluation, I created a dummy variable to identify which template was given to each participant (0 = No, 1 = Yes). These were used as covariates in the hypothesis tests.

Participants responded to the email and provided the requested information. They received a thank you reply from the designer and were told that the designer would deliver their new design within 24 hrs, per the Netwrk service policy. The following day, approximately one hour after the end of the 24-hour window, participants received a third email from the designer with the personalized logo design (see Appendix I for examples). The email included a link to the second survey, on which participants could evaluate the service experience. On the survey, participants were asked to provide a rating on a 5-star scale. They were told that the rating would be assigned to the designer, and would be shared with Netwrk so that they could post it on their website. Participants also indicated *expectancy disconfirmation* (1 item), *satisfaction* (1 item), and *trust* (four items; α = .92).

6.5   Study 2B Results

6.5.1   Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 2B are in Appendix J.

Descriptive statistics by condition are in Table 8. As expected, uncertainty was higher in

the peer-to-peer (M = 3.17) compared to commercial condition (M = 2.86 ; p = .05), and

uncertainty did not differ between the risk level conditions (M_{High} = 3.04; M_{Low} = 2.99; p

= .75). However, the risk level manipulation did not significantly increase perceived risk

in the high (M = 3.01) compared to the low risk condition (M = 2.83 ; p = .37). Perceived

risk also did not significantly differ between the peer-to-peer (M = 2.87) and commercial

conditions (M = 2.97 ; p = .62).

Table 8: Means and Standard Deviations for Study 2B

	P2P High Risk N = 46	P2P Low Risk N = 44	Commercial High Risk N = 47	Commercial Low Risk N = 48
Expectations	5.02 (1.33)	5.05 (1.16)	4.89 (1.15)	5.25 (1.06)
Perceived Risk	2.98 (1.31)	2.75 (1.51)	3.04 (1.32)	2.90 (1.43)
Uncertainty	3.20 (1.20)	3.15 (1.02)	2.88 (1.00)	2.84 (.95)
Rating	4.07 (.95)	4.20 (.93)	3.96 (1.12)	4.21 (.80)
Expectancy Disc.	4.70 (1.64)	4.57 (1.50)	4.51 (1.57)	4.35 (1.41)
Satisfaction	5.11 (1.55)	5.43 (1.50)	5.04 (1.49)	5.15 (1.38)
Trust	5.74 (1.14)	5.82 (1.37)	5.72 (1.24)	5.67 (.97)

6.5.2 Tests of Hypotheses

Next, hypotheses $H_{2.3}$ and $H_{2.4}$ were tested. Supporting $H_{2.3}$, there was a significant

positive relationship between trust and rating (β = .58, SE = .04, p < .01). To support $H_{2.4}$,

I expected to find a three-way interaction such that the effect of provider trust on rating

was highest for the high risk peer-to-peer condition, because of higher perceived risk in that condition. From a moderated moderation analysis (PROCESS model 3; Hayes 2018), there was a significant 2-way interaction such that the effect of trust on rating was stronger in the high risk condition ($\beta = .18$, SE $= .09$, $p = .03$). However, the 2-way interaction with service type was not significant ($\beta = .16$, SE $= .20$, $p = .42$) and the three-way interaction was also not significant ($\beta = -.11$, SE $= .17$, $p = .52$). Next, a moderation analysis was conducted to directly test H$_{2.4}$ (PROCESS model 1; Hayes 2018). Trust was the predictor, rating was the dependent variable, and perceived risk was the moderator. However, the interaction between trust and perceived risk was not significant ($\beta = .01$, SE $= .03$, $p = .63$).

Finally, to test H$_{2.5}$, a multiple moderated mediation analysis was conducted (PROCESS model 21; Hayes 2018). Expectancy disconfirmation was the predictor, trust was the mediator, and rating was the dependent variable. Uncertainty was the first moderator, and perceived risk was the second moderator. There was a significant positive main effect of expectancy disconfirmation on trust (which supports H$_{2.1}$; $\beta = .50$, SE $= .04$, $p < .01$). There was also a significant negative main effect of uncertainty on trust ($\beta = -.30$, SE $= .06$, $p < .01$). As expected, there was a significant positive interaction between expectancy disconfirmation and uncertainty (supporting H$_{2.2}$; $\beta = .06$, SE $= .03$, $p < .05$, d $= .31$) such that the effect of expectancy disconfirmation on trust was stronger when uncertainty was higher. The effect size of this interaction was larger in Study 2B than Study 2A, and the interaction was significant, which might be related to the fact that participants were removed if they did not believe the deception. Further, there was a

positive main effect of trust on rating (which supports $H_{2.3}$; $\beta = .33$, SE $= .05$, $p < .01$). The main effect of perceived risk on rating was not significant ($\beta = -.03$, SE $= .03$, $p = .45$), and the interaction between trust and perceived risk was also not significant ($\beta = .01$, SE $= .03$, $p = .63$). Further, there was a significant direct effect of expectancy disconfirmation on rating ($\beta = .29$, SE $= .04$, $p < .01$). The indirect effect of expectancy disconfirmation through the mediator trust was significant at all levels of the moderators.

Finally, the full model from Figure 5 was tested to demonstrate that trust has an effect on ratings beyond the effect of satisfaction (as described in P_4). Perceived risk was not included in the analysis because of the earlier non-significant result. A parallel moderated mediation analysis was conducted with expectancy disconfirmation as the predictor, satisfaction and trust as parallel mediators, uncertainty as the moderator, and rating as the dependent variable (PROCESS model 7; Hayes 2018). That is, the path from expectancy disconfirmation to *satisfaction* to rating (moderated by uncertainty) was tested in parallel to the path from expectancy disconfirmation to *trust* to rating (see Figure 6). As expected, the interaction between expectancy disconfirmation and uncertainty on satisfaction was not significant ($\beta = .02$, SE $= .04$, p $= .63$). Also as expected, the indirect effect of expectancy disconfirmation on rating through *satisfaction* was significant at all levels of uncertainty, and the index of moderated mediation for this path was not significant (IMM $= .01$, 95% CI: -.02, .03). The indirect effect of expectancy disconfirmation on rating through *trust* was also significant at all levels of uncertainty. However, providing some support for the proposed process, the index of moderated mediation was marginally significant (i.e. 90% confidence; IMM $= .01$, 95% CI: .001, .02). The effect size of the

indirect effect through trust was larger when uncertainty was *high* (M_{+1SD} = 4.06;

βindirect = .06, 95% CI: .001, .13), compared to when uncertainty was *medium* (M_{MEAN} =

3.01; βindirect = .05, 95% CI: .001, .11) and when uncertainty was *low* (M_{-1SD} = 1.97;

βindirect =.04, 95% CI: .001, .10). Finally, there was a significant direct effect of

expectancy disconfirmation on rating (β = .15, SE = .04, p < .01).
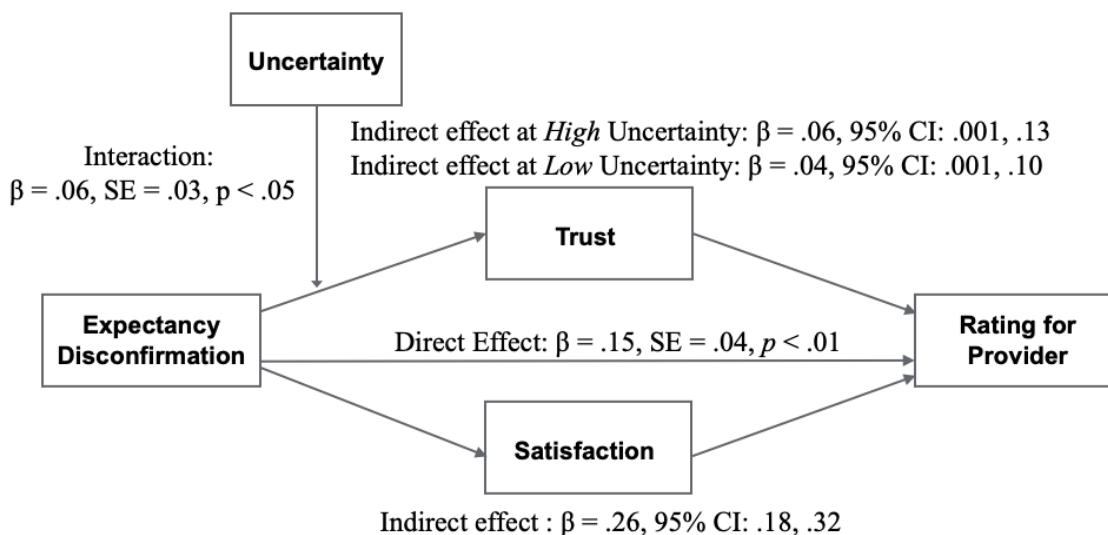
Figure 6: Results of Study 2B Parallel Moderated Mediation Analysis

6.6   Discussion and Subsequent Analysis

Study 2B provides support that expectancy disconfirmation leads to trust, and that this

relationship is strengthened when uncertainty is high. To improve the precision of the

results in this study, I measured and removed participants who suspected that Netwrk was

not a real company. However, it is possible that the participants who suspected the

deception may, in general, have a lower overall level of trust. Thus their removal may be

confounded with our dependent variable of trust in this analysis. If I include these

participants in the moderation analysis to test $H_{2.2}$ (PROCESS model 1; Hayes 2018), the predicted interaction between expectancy disconfirmation and uncertainty on trust is no longer significant ($\beta = .04$, SE $= .03$, $p = .15$). I still believe that it is more appropriate to remove these participants because if they did not believe that they were transacting with a real designer and a real company, then they would not feel the same level of trust or distrust as they would if they knew that Netwrk was fake. However, for robustness, I will test this relationship again in Study 3 with customers of a real peer-to-peer company who report on their actual perceptions of trust in their providers. I also note that the removal of the participants who suspected the deception did not materially affect any of the other reported results in Study 2B.

The manipulation for perceived risk failed to generate significant differences between conditions, and perceived risk also did not significantly differ between the peer-to-peer and the commercial conditions. These results may have been impacted by the fact that perceived risk was relatively low in all conditions. In retrospect, the resume design experience might have worked better in Study 2B because resumes include more personal information and thus may be considered to be riskier. The logo design experience might have worked better in Study 2A because logos have more uncertainty than resumes (which are built from existing versions and are more standardized in format).

Study 2B also finds support that trust is positively related to ratings. However, the predicted moderation by perceived risk on the relationship between trust and ratings ($H_{2.4}$) was not significant. This result raises some questions about the conceptual model in

terms of the ways in which perceptions of trust affect ratings. In particular, upon reflection, the effect of trust on ratings may be partly realized through increased satisfaction (i.e., trust → satisfaction → ratings) and this relationship may be affected by perceived risk. Satisfaction and trust are closely related constructs, and both have been found to be associated with word-of-mouth (Ranaweera & Prabhu, 2003). Both constructs are important factors in exchange relationships (Garbarino & Johnson, 1999; Smith, 1998), but there has been relatively little research devoted to exploring the relationship between them (Selnes, 1998). While many researchers have proposed a causal relationship between satisfaction and trust, some have argued that satisfaction leads to trust (e.g., Ganesan, 1994; Martin et al., 2011; Singh & Sirdeshmukh, 2000) while others have argued that trust leads to satisfaction (e.g., Chen & Chou, 2012; Chiou & Droge, 2006; Smith & Barclay, 1997).

The direction of causality between trust and satisfaction may be context dependent, but the moderators of the relationship have not been rigorously explored. One factor that may affect the relationship is perceived risk. Indeed, the role of trust is to reduce perceived risk, which can lead to satisfaction (Chen & Chou, 2012). Thus if perceived risk is high before an exchange, it stands to reason that demonstrated trustworthiness in this high risk environment should drive higher levels of satisfaction. The argument for this relationship is similar to my earlier argument about why perceived risk should strengthen the relationship between trust and ratings (i.e., $H_{2.4}$). When risk is high and the outcome of service failure is severe (e.g., a bungee jumping experience), then provider trustworthiness becomes more important. Demonstrated trustworthiness, through

expectancy disconfirmation, should therefore have a stronger impact on satisfaction when risk is high, above and beyond other important factors such as product quality.

To test this relationship, a moderation analysis was conducted with trust as the predictor, satisfaction as the dependent variable, and perceived risk as the moderator (PROCESS model 1; Hayes 2018). There was a significant positive effect of trust on satisfaction (β = .98, SE = .06, p < .01) and as predicted, the interaction was positive and significant (β = .09, SE = .04, p = .02) such that the effect of trust on satisfaction is strengthened when perceived risk is higher. Further, when expectancy disconfirmation and uncertainty were added to the model (i.e., PROCESS model 21; Hayes 2018) we find the two expected interactions; there was a significant positive interaction between expectancy disconfirmation and uncertainty on trust (β = .06, SE = .03, p = .04) and a significant positive interaction between trust and perceived risk on satisfaction (β = .09, SE = .03, p = .01). Thus, the subsequent analysis provides some support that trust may lead to increased satisfaction when perceived risk is high. These relationships will be tested again in Study 3, which will also test the impact of ratings importance and need to justify on the effects of satisfaction and trust on ratings.

7 Study 3

Study 2 demonstrated that both trust and satisfaction are important for service evaluations when uncertainty and perceived risk are relatively high, as they are in peer-to-peer services. The objective of Study 3 is to begin to understand how the effects of trust and

satisfaction lead to positive bias in peer-to-peer ratings through propositions P_5, P_{6a} and P_{6b}. Specifically, I expect that when expectations are merely met (rather than exceeded), trust is higher than satisfaction (P_5). This is because to demonstrate trustworthiness, providers are only required to meet (and not exceed) their promises and commitments. Thus, when expectations are merely met, ratings that are more affected by trust will be higher than ratings that are more affected by satisfaction.

But what causes the effect of satisfaction to be reduced relative to the effect of trust? I propose that network factors in peer-to-peer platforms, namely, a feeling that ratings are more important to peer providers, and that ratings need to be justified, weakens the effect of satisfaction on ratings (i.e., P_{6a}). Satisfaction is based on fulfillment of needs, and needs are unique to the individual. Thus satisfaction is somewhat subjective and may be difficult to justify. On the other hand, trust is demonstrated when a provider meets commitments (i.e., when expectations are confirmed). If a provider does not meet specific commitments, lower ratings are easier to justify. But if a provider meets commitments, it would be difficult to justify a lower rating, especially because ratings are so important for peer providers. The importance of ratings may further contribute to bias by motivating consumers to post higher ratings than may otherwise be deserved (i.e., P_{6b}). As long as the provider demonstrated trustworthiness, consumers who recognize the importance of ratings may post higher ratings so that they don't harm providers from future business.

Study 1 and Study 2 were performed in a controlled environment that was not able to mimic the network effects of real peer-to-peer platforms such as Airbnb. Study 3 will

address these concerns through a field experiment with an actual peer-to-peer service. I

further manipulate the perceived anonymity of ratings and the ratings visibility (public or

private) to determine whether they can reduce perceived ratings importance and a need to

justify ratings. In doing so, I can test whether positive ratings bias can be attenuated

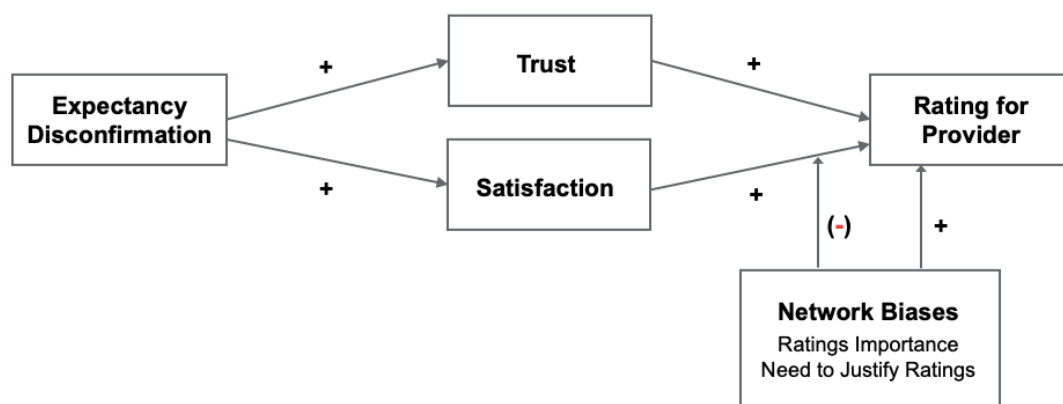through these changes. Figure 7 illustrates the model to be tested in this study.

Figure 7: Model to be tested in Study 3

Finally, Study 3 also tests two new hypotheses that stem from the subsequent analysis

performed in Study 2B. That is, that trust is positively related to satisfaction, and that

perceived risk moderates the relationship between trust and satisfaction. Table 9 lists the

hypotheses to be tested in Study 3.

Table 9: Study 3 Hypotheses

3.1	Satisfaction is positively related to ratings.
3.2	A need to justify ratings decisions moderates the relationship between satisfaction and ratings; this relationship will be weaker when need to justify is higher.
3.3	Perceived ratings importance moderates the relationship between satisfaction and consumer ratings; this relationship will be weaker when ratings importance is higher.
3.4	Perceived ratings importance is positively related to ratings.
3.5	Trust is positively related to satisfaction.
3.6	Perceived risk moderates the relationship between trust and satisfaction; this relationship will be stronger when perceived risk is higher.

7.1   Method

7.1.1   Participants & Design.

Three hundred and seventy-one participants were recruited for a field experiment with RVezy, a peer-to-peer platform for recreational vehicles. Participants were randomly assigned to one of four conditions in a 2 (*Rating Visibility*: public vs. private) x 2 (*Rater Identification*: individual vs. anonymous) between-subjects design. Participants were customers of RVezy and completed two surveys. The first survey was completed shortly after booking the RV rental but *prior* to taking possession of the vehicle, and the second was completed shortly *after* the rental experience had ended. Thirteen participants (3.5%) were removed from the study because of inconsistent results, likely stemming from a survey visibility issue on mobile devices that caused some answer choices to not appear on screen. On a single-item question, these thirteen participants rated provider trust at the top of the seven-point scale (which was similar to their ratings for satisfaction and which

matched their feelings toward the provider in an open-ended question), but then they rated multiple-item reliability and integrity (and three other items on the same measure) at the bottom of the scale (i.e., 1 of 7). The top-of-scale answer choices for these seven items did not appear on screen on mobile devices which likely caused this issue. The final sample included 358 RVezy customers (196 women; Mage = 46.44 years). The majority of participants (91.3%) had no prior rental experience with RVezy, while 7.3% had rented from RVezy one other time, and the remaining 1.4% had rented from RVezy two or more times in the past.

The field experiment was conducted during the prime rental season from June to Sept. 2019. The RV rental process begins online. Consumers who navigate to the RVezy.com website can view rental prices, names and photos of the owners, details about the vehicle (which are provided by the owners), and ratings and reviews from prior customers (see Appendix K). Bookings are made through the website. Immediately after finalizing a booking during the study period, customers received an email from RVezy asking if they would like to take part in the study in exchange for a chance to win an iPad prize. The email contained a link to the first survey, where participants indicated their *expectations*, *perceived risk*, and *uncertainty* (α = .87). These variables were measured the same as in the prior studies. The correlation between perceived risk and uncertainty was moderate (r = .46, p < .01). Participants completed the survey, and then began their rental experience.

After the experience, at the close of rental, the participant received an email from RVezy with a link to the second survey. The survey stated "Before completing the survey, please

rate your RV experience. Specifically, please rate the RV and RV owner on a five-star scale". The instructions were designed to make the participant believe that their ratings were part of the real RVezy ratings process, rather than being delivered only as part of the research. Additional instructions were included to operationalize the two factors. In the *public rating conditions*, participants were told that their rating would be posted publicly to the RVezy.com website on the profile page for the RV owner that they rented from. They were further told that their feedback was important for other RVezy consumers to help them decide which RV to choose for their rental. I expected that these instructions would prompt participants to consider the importance of ratings to providers for attracting future customers. In the *private rating conditions*, participants were told that their rating would be posted privately to the RVezy owner. It would not be posted to the website and would not be viewable by other consumers. They were further told that their feedback was important for RVezy owners to help them understand how well they are performing. I expected that, compared to the public conditions, these instructions would make participants feel that their rating was relatively less important. This is because private ratings, which are not visible to other consumers, cannot affect the provider's reputation and future business with those consumers. I also expected that because the ratings were sent directly to the provider, participants may still feel strongly that they might have to justify their ratings decision to the provider.

Further, in the *individual rating conditions*, participants were told that their rating represented their individual feedback. I expected that these instructions would prompt participants to consider the fact that the provider will know exactly what rating the

participant gave. In the *anonymous rating conditions*, participants were told that the RV

owner would not see their individual rating because it would be aggregated (i.e.,

"averaged") with the ratings from other consumers who had previously rented the same

RV. I expected that, in comparison to the individual rating conditions, these instructions

should make participants feel that their rating is less important. This is because their

rating is part of collective feedback, and may not dramatically affect the overall aggregate

score for the provider. I also expected that need to justify should be lower because

providers cannot identify the participant's specific rating. Providers should thus be less

likely to question the rating or to retaliate against a low rating.

Next, participants indicated whether, as they were making their rating decision, they felt a

need to justify the rating. This was a five-item measure ($\alpha = .73$). Participants were given

the instruction "When I was making my ratings decision…" ("I felt that I might need to

justify my ratings choice to the RV owner," "I felt that I might have to explain my ratings

choice," "I felt that my rating needed to be completely fair to the RV owner," "I felt that I

needed to have clear reasons to support my ratings choice," "I felt that I might be

criticized by the RV owner for my ratings choice;" 1 = Not at all, 7 = Very Much). Next,

ratings importance (3 items; $\alpha = .80$) was measured similarly to the Pretest Study ("How

important is your rating to your RV owner," "How closely do you think your RV owner

will monitor your online rating," "How much do you think your rating will affect the RV

owner's future rentals;" 1 = Not at all, 7 = Very/Very Much).

Participants next rated their *expectancy disconfirmation* (1 item) and *satisfaction* (1 item)*,*

which were measured the same as in Study 2. *Trust* was measured by the same four items

as in previous studies ($\alpha = .95$) but was rated on a seven-point semantic differential scale

("My provider has very low integrity/My provider has very high integrity," "My provider

definitely cannot be counted on to do what's right/My provider definitely can be counted

on to do what's right," "My provider is very unreliable/My provider is very reliable,"

"My provider is very incompetent/My provider is very competent"). Three additional

items were included on the same semantic differential scale to gather additional feedback

about the provider: *helpfulness* ("My provider is very unhelpful/My provider is very

helpful"), *friendliness* ("My provider is very unfriendly/My provider is very friendly"),

and *professionalism* ("My provider is very unprofessional/My provider is very

professional"). An additional one-item measure for provider *trustworthiness* was included

to determine how well that the four-item measure (which taps into the two dimensions of

integrity and reliability) would correlate with this more global measure ("How

trustworthy is the RV owner that you rented from?; 1 = Very Untrustworthy, 7 = Very

Trustworthy"). Results confirmed that the two measures were moderately to strongly

correlated ($r = .69$, $p < .01$). Finally, participants answered an open-ended question to

describe how they made their ratings decision.

7.2   Results

7.2.1   Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 3 are in Appendix L.

Descriptive statistics by condition are in Table 10. I expected that both ratings importance

and need to justify ratings would be higher in the individual vs. anonymous condition,

and that ratings importance would also be higher in the public vs. private ratings

condition. As expected, ratings were considered to be more important in the individual

(M = 6.18) compared to the anonymous (i.e., aggregated rating) condition (M = 5.97, F(1,

355) = 4.14, p < .05). However, there was no significant difference in need to justify

between the two conditions ($M_{Individual}$ = 3.68, $M_{Anonymous}$ = 3.74, F(1, 355) = 1.40, p =

.75). This may have been due to confusion with how the question was interpreted.

Although the instructions for this question asked participants to reflect on whether they

considered the need to justify while they were *making their decision* on ratings, the rating

itself may have impacted feelings toward the need to justify *after* the decision. That is, for

some participants, rather than a need to justify causing them to artificially inflate their

ratings (i.e., make the ratings more positive so that they won't have to justify them to the

provider), the fact that the rating was high caused them to answer that they felt less likely

that they *then* needed to justify the rating. Thus, rather than a positive relationship

between need to justify and ratings (i.e., a high need to justify leads to high ratings), I

found a negative relationship in this data (i.e., a high rating leads to lower feelings that

the rating would then need to be justified; β = -.40, SE = .11, p < .01).

As expected, the need to justify also did not differ between the public and private

conditions (M_{Public} = 3.61, $M_{Private}$ = 3.81, F(1, 355) = 1.91, p = .17). However,

unexpectedly, there was no significant difference in ratings importance for these

conditions (M_{Public} = 6.08, $M_{Private}$ = 6.06, F(1, 355) = .06, p = .80). Planned contrasts did

not show any difference in ratings importance between the individual public (M = 6.25)

and individual private conditions (M = 6.11, t(353) = .96, *p* = .34), nor any difference

between the anonymous public (M = 5.93) and anonymous private conditions (M = 6.01,

t(353) = -.49, *p* = .63). Thus the public and private conditions were collapsed into the

individual and anonymous conditions for the rest of the analyses.

Table 10: Means and Standard Deviations for Study 3

	Individual Public N = 86	Individual Private N = 90	Anonymous Public N = 95	Anonymous Private N = 87
Expectations	5.92 (.96)	6.02 (.82)	5.99 (.98)	6.15 (1.03)
Perceived Risk	3.50 (1.47)	3.21 (1.47)	3.38 (1.63)	3.05 (1.50)
Uncertainty	2.48 (.96)	2.17 (.95)	2.42 (1.08)	2.19 1.08)
Rating	4.78 (.52)	4.73 (.58)	4.52 (.90)	4.63 (.63)
Need to Justify	3.52 (1.25)	3.84 (1.38)	3.70 (1.35)	3.78 (1.51)
Ratings Importance	6.25 (.80)	6.11 (.99)	5.93 (1.06)	6.01 (1.03)
Expectancy Disc.	5.97 (1.24)	6.03 (1.17)	5.56 (1.49)	5.76 (1.23)
Satisfaction	6.47 (.94)	6.50 (.72)	6.15 (1.37)	6.44 (.87)
Trust	6.79 (.49)	6.83 (.45)	6.48 (1.04)	6.75 (.64)
Helpful	6.87 (.43)	6.86 (.57)	6.57 (1.11)	6.82 (.54)
Friendly	6.91 (.33)	6.93 (.36)	6.67 (.94)	6.84 (.55)
Professional	6.74 (.67)	6.73 (.72)	6.39 (1.17)	6.69 (.78)
Trustworthiness	6.83 (.47)	6.81 (.67)	6.58 (.94)	6.72 (.64)

I next compared the individual conditions to the anonymous conditions for the other

variables. As expected, the manipulations did not affect participants' expectations

($M_{Individual}$ = 5.97, $M_{Anonymous}$ = 6.07, F(1, 356) = .88, *p* = .35), perceived risk ($M_{Individual}$ =

3.35, $M_{Anonymous} = 3.22$, $F(1, 356) = .68$, $p = .41$), and uncertainty ($M_{Individual} = 2.32$, $M_{Anonymous} = 2.31$, $F(1, 356) = .01$, $p = .94$).

As expected, ratings were higher in the individual conditions ($M = 4.76$) compared to the anonymous conditions ($M = 4.57$, $F(1, 356) = 6.63$, $p < .05$). However, unexpectedly, expectancy disconfirmation ($M_{Individual} = 6.00$, $M_{Anonymous} = 5.66$, $F(1, 356) = 6.39$, $p < .05$) and trust ($M_{Individual} = 6.81$, $M_{Anonymous} = 6.61$, $F(1, 356) = 7.39$, $p < .01$) were also significantly higher in the individual conditions, and satisfaction was marginally higher as well ($M_{Individual} = 6.48$, $M_{Anonymous} = 6.29$, $F(1, 356) = 3.38$, $p < .10$). Consumers in the individual condition also considered their providers to be significantly more helpful ($M_{Individual} = 6.86$, $M_{Anonymous} = 6.69$, $F(1, 356) = 5.30$, $p < .05$), friendly ($M_{Individual} = 6.92$, $M_{Anonymous} = 6.75$, $F(1, 356) = 6.85$, $p < .01$), and professional ($M_{Individual} = 6.74$, $M_{Anonymous} = 6.53$, $F(1, 356) = 5.00$, $p < .05$). It's possible that the actual quality of the vehicles and providers in the individual conditions might have been significantly higher on average than in the anonymous conditions, but this is unlikely given the sample size. Instead it is more likely that, after providing their rating, some participants may have felt the need to seek consistency between the rating and their evaluations so that the rating did not appear to be or feel to be biased. Participants may also have felt that their answers to these questions might be shared with the RV owner in the individual conditions, which may have caused them to be assessed higher.

7.2.2   Tests of Proposition P₅ and Hypotheses

Next, P_5 was tested. Proposition P_5 states that trust will be higher than satisfaction when expectations are merely met. This proposition is not directly testable because it requires the statistical comparison of two different variables. However, both variables were similarly operationalized as seven-point semantic differential scales, so I performed a t-test comparison of their mean scores for all participants who indicated that their expectations were merely met (N = 39). Supporting P_5, the difference between the mean score for satisfaction (M = 6.13) and the mean score for trust (M = 6.62) was statistically significant (t(76) = -2.59, $p < .05$). Trust was significantly higher than satisfaction when expectations were merely met.

Next, the hypotheses were tested. To test $H_{3.1}$ and $H_{3.2}$, a moderation analysis was conducted with satisfaction as the predictor, rating as the dependent variable, and need to justify as the moderator. As expected, the main effect of need to justify on ratings was not significant ($\beta = -.02$, SE = .02, p = .31). Supporting $H_{3.1}$, there was a significant positive main effect of satisfaction on rating ($\beta = .53$, SE = .03, p < .01). However, the interaction was not significant ($\beta = -.02$, SE = .02, p = .37) which again may be due to issues with how the question for need to justify was interpreted.

Next, to test $H_{3.1}$, $H_{3.3}$ and $H_{3.4}$, a moderation analysis was conducted with satisfaction as the predictor, rating as the dependent variable, and ratings importance as the moderator (PROCESS model 1; Hayes 2018). Supporting $H_{3.1}$, there was a significant positive main effect of satisfaction on rating ($\beta = .36$, SE = .03, p < .01). Supporting $H_{3.4}$, there was a significant positive main effect of ratings importance ($\beta = .12$, SE = .02, p < .01). Finally,

supporting H$_{3.3}$, there was a significant negative interaction (β = -.13, SE = .02, p < .01, d = .70) such that the effect of satisfaction on rating is weaker when ratings importance is higher.

In a separate test of these hypotheses, I used the rater identification conditions as a proxy measure for ratings importance (recall that ratings importance is higher when a rater believes that their rating is identifiable). First, a moderation analysis was conducted with satisfaction as the predictor, rating as the dependent variable, and rater identification as the moderator. Rater identification was a dummy variable (0 = anonymous, 1 = individual). Supporting H$_{3.1}$, there was a significant positive main effect of satisfaction on rating (β = .50, SE = .02, p < .01). Supporting H$_{3.4}$, there was a significant positive main effect of rater identification, such that ratings are higher when raters believe that they can be individually identified by providers (β = .09, SE = .05, p < .01). Finally, supporting H$_{3.3}$, there was a significant negative interaction (β = -.10, SE = .05, p < .05, d = .23) such that the effect of satisfaction on rating is weaker when raters are individually identified compared to when they are anonymous.

To provide further support for P$_5$ and P$_6$, I tested the full indirect relationship from expectancy disconfirmation to ratings using a series of parallel moderated mediation analyses (PROCESS model 14; Hayes 2018). The indirect path from expectancy disconfirmation to ratings through the mediator satisfaction was tested in parallel to the indirect path through the mediator trust. The analyses separately tested the three different moderators for these mediations (i.e., need to justify, ratings importance, and rater

identification). First, *need to justify* was the moderator. There was a significant positive main effect of expectancy disconfirmation on rating ($\beta = .13$, SE $= .02$, $p < .01$), and as expected, there was a significant main effect of trust on rating ($\beta = .12$, SE $= .04$, $p = .01$) and satisfaction on rating ($\beta = .36$, SE $= .03$, $p < .01$). As expected, the main effect of need to justify was not significant ($\beta = .01$, SE $= .02$, $p = .98$). Supporting $H_{3.2}$, there was a significant *negative* interaction between satisfaction and need to justify ($\beta = -.07$, SE $= .02$, $p < .01$, d $= .18$) such that the effect of satisfaction on rating is weaker when a consumer feels a higher need to justify their rating. The indirect effect of disconfirmation on ratings through the mediator satisfaction was significant for all levels of need to justify, but the effect size was smaller for those who felt a relatively *high* need to justify ($M_{+1SD} = 5.09$; $\beta_{indirect} = .14$, 95% CI: .08, .22) compared to those who felt a relatively *low* need to justify ($M_{-1SD} = 2.33$; $\beta_{indirect} = .24$, 95% CI: .14, .32). I expected that need to justify would not similarly reduce the effect of trust on ratings, because trust assessments (based on whether the provider met commitments) are easier to justify. However, in fact, need to justify further strengthened the effect of trust. There was a significant *positive* interaction between trust and need to justify ($\beta = .09$, SE $= .03$, $p < .01$, d $= .14$) such that the effect of trust on rating is stronger when a consumer feels a higher need to justify their rating (see Figure 8). A Johnson-Neyman analysis showed that this indirect effect was significant when need to justify was at a value of 3.40 (out of 7) or higher (approximately 45% of participants). Further, the indirect effect of disconfirmation on ratings through the mediator trust was significant for individuals who felt a relatively *higher* need to justify ($M_{+1SD} = 5.09$; $\beta_{indirect} = .07$, 95% CI: .02, .12) or *medium* need to justify ($M_{MEAN} = 3.71$; $\beta_{indirect} = .04$, 95% CI: .002, .11), but the indirect effect was not

significant for those who felt a relatively *lower* need to justify (M_{-1SD} = 2.33; βindirect =.00, 95% CI: -.05, .12).
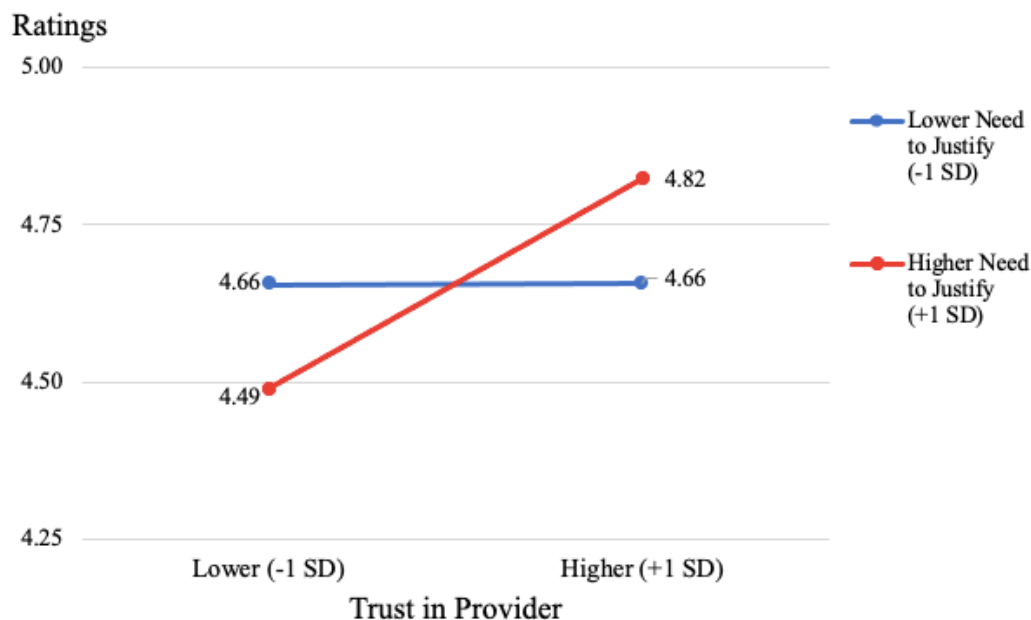

Figure 8: The Effect of Trust on Ratings at Different Levels of Need to Justify

Next, *ratings importance* was included in the analysis as the moderator for the relationship between trust and ratings and between satisfaction and ratings. There was a significant positive main effect of disconfirmation on rating (β = .12, SE = .02, p < .01). Supporting H$_{3.1}$, there was a significant positive main effect of satisfaction on rating (β = .23, SE = .04, p < .01) and supporting H$_{3.4}$, there was a significant positive main effect of ratings importance on rating (β = .10, SE = .02, p < .01). The main effect of trust was not significant (β = .07, SE = .06, p = .25). As expected, the interaction between trust and ratings importance was also not significant (β = .01, SE = .04, p = .87). Finally, supporting H$_{3.3}$, there was a significant negative interaction between satisfaction and

ratings importance (β =-.13, SE = .03, p < .01, d = .71) such that the effect of satisfaction

on rating is weaker when a consumer feels that their individual rating is more important

to the provider (see Figure 9). Further, the indirect effect of disconfirmation on ratings

through satisfaction was significant for individuals who felt a relatively *lower* ratings

importance (M_{-1SD} = 5.09; βindirect = .19, 95% CI: .13, .25) or *medium* ratings

importance (M_{MEAN} = 6.07; βindirect = .12, 95% CI: .08, .17). As expected, the indirect

effect was not significant for those who felt a *higher* ratings importance (M_{+1SD} = 7.00;

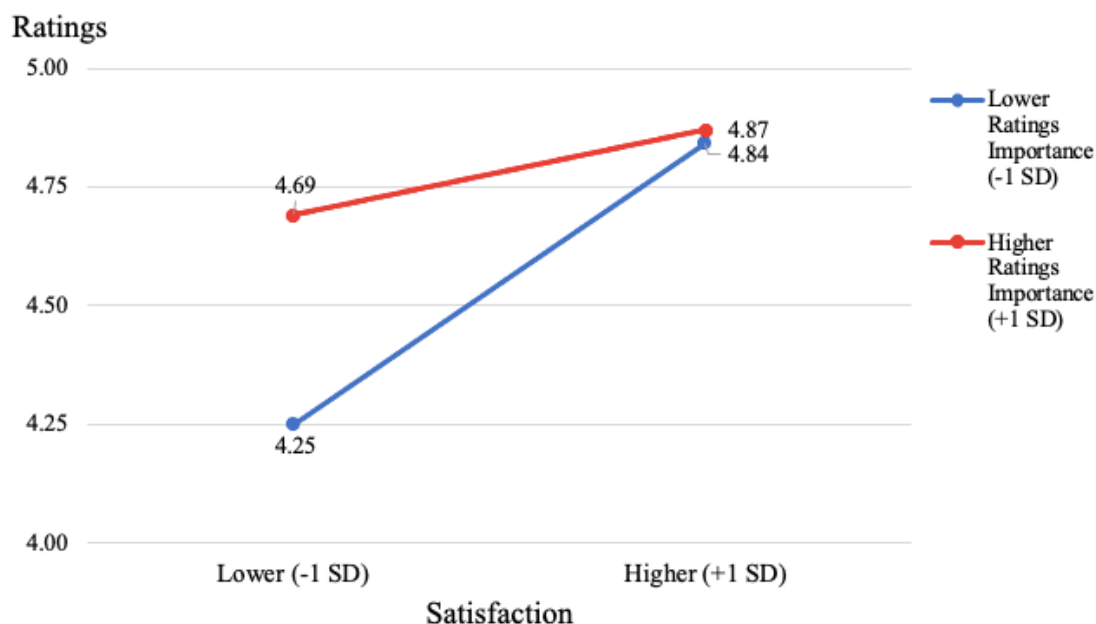βindirect =.06, 95% CI: -.01, .12).
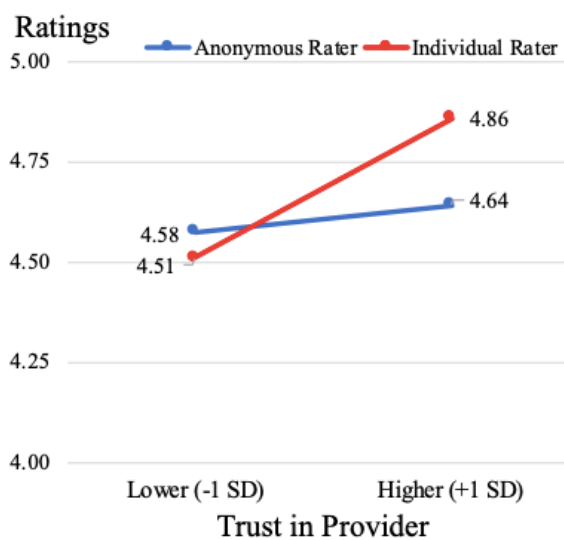

Figure 9: The Effect of Satisfaction on Ratings at Different Levels of Ratings Importance

Finally, the same analysis was conducted, but *rater identification* was the moderator.

There was a significant positive main effect of disconfirmation on rating (β = .13, SE =

.02, p < .01). As expected, there was also a significant main effect of trust on rating (β =

.09, SE = .05, p = .05) and satisfaction on rating (β = .40, SE = .04, $p < .01$). However, the main effect of rater identification was not significant (β = .11, SE = .50, p = .83). As expected, the interaction between trust and rater identification was also not significant (β = .15, SE = .10, p = .13). Supporting H$_{3.3}$, there was a significant negative interaction between satisfaction and rater identification (β =-.17, SE = .06, $p < .01$, d = .25) such that the effect of satisfaction on rating is weaker when a consumer feels that their rating will be individually identifiable (see Figure 10). The indirect effect of expectancy disconfirmation on ratings through satisfaction was significant in both conditions, but the effect size was larger in the *anonymous* (βindirect$_{anonymous}$ = .21, 95% CI: .13, .29) compared to *individual* condition (βindirect$_{individual}$ = .12, 95% CI: .06, .19).
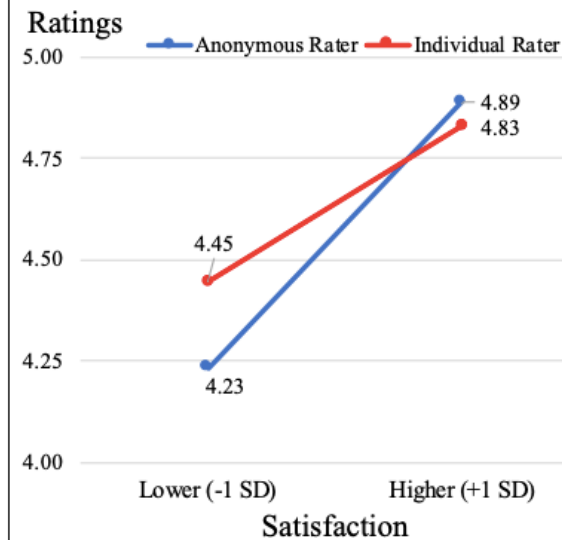

Figure 10: Moderating Effects of Anonymous vs. Identified Rater Conditions on the Effect of Trust and Satisfaction on Ratings

7.2.3   Tests of Post Hoc Hypotheses from Study 2B

Finally, the post hoc hypotheses from Study 2B ($H_{3.5}$ and $H_{3.6}$) were tested. In Study 2B, results of a subsequent analysis suggested that trust may lead to satisfaction, and that the relationship is stronger when perceived risk is high. To test these hypotheses, I conducted a multiple moderated mediation analysis (PROCESS model 21; Hayes 2018). Expectancy disconfirmation was the predictor, trust was the mediator, and satisfaction was the dependent variable. Uncertainty was the first moderator, and perceived risk was the second moderator. As expected, there was a significant positive main effect of expectancy disconfirmation on trust ($\beta = .28$, SE $= .03$, $p < .01$), and a negative main effect of uncertainty on trust ($\beta = -.07$, SE $= .03$, $p = .03$). As expected, there was also a significant positive interaction between expectancy disconfirmation and uncertainty ($\beta = .06$, SE $= .02$, $p = .01$, d $= .31$) such that the effect of expectancy disconfirmation on trust was stronger when uncertainty was higher. Further, as expected, there was a significant positive main effect of expectancy disconfirmation on satisfaction ($\beta = .34$, SE $= .03$, $p < .01$) and the main effect of perceived risk on satisfaction was not significant ($\beta = -.01$, SE $= .02$, $p = .56$). Supporting $H_{3.5}$, there was a positive main effect of trust on satisfaction ($\beta = .65$, SE $= .06$, p $< .01$). However, the predicted interaction between trust and perceived risk on satisfaction was not significant ($\beta = .04$, SE $= .02$, $p = .12$). Future research should continue to test this new hypothesis.

7.3   Discussion

Study 3 provides support that contextual network-related factors in peer-to-peer

exchanges affect the relationship between expectancy disconfirmation and ratings.

Ratings importance and the need to justify ratings (which were shown in the Pretest

Study to be higher in peer-to-peer services), decrease the importance of satisfaction at the

expense of trust (supporting P_{6a}). Performance that merely meets (rather than exceeds)

expectations generates only moderate levels of satisfaction. But peer-to-peer consumers

will give high ratings to providers even if expectations are merely met. This is because

although consumers may not be fully satisfied, providers have demonstrated that they can

be trusted. Thus it would be difficult to justify giving a lower rating. Indeed, I had

expected that need to justify would not affect trust, but the results showed that when need

to justify is higher, the effect on trust on rating is actually stronger, while the effect of

satisfaction is weaker. Further, the results showed that ratings importance creates an

additional bias through a direct positive effect on peer-to-peer ratings (supporting P_{6b}).

Higher levels of perceived importance lead to higher ratings, likely because consumers

don't want to harm providers from future business.

Study 3 also provides a possible solution to help attenuate the positive bias. Average

ratings were lower if consumers felt that their rating was anonymous. This is because

consumers in the anonymous (vs. individual) condition felt that their ratings were

relatively less important to the peer provider. For these consumers, satisfaction was more

strongly related to ratings, which further suggests that these ratings were less biased.

Some platforms already attempt to anonymize ratings, but platforms could do more to

help consumers believe that providers will not be able to identify their rating and not be

able to respond with retribution. I had also expected that ratings would be considered to be relatively less important in the private (vs. public) condition, but there were no significant differences. This result might be because participants in the private condition knew that both the provider and RVezy would see the rating. RVezy could therefore use the rating for evaluative purposes, and hence the rating would still be important to the provider's future business. Participants may have also thought that the rating would eventually be placed on the RVezy website, because they likely saw ratings from other customers on the website when they were booking their rental.

Study 3 demonstrated that even when expectations are merely met, they may lead to higher ratings for peer-to-peer services. The remaining studies attempt to show that consumers of peer-to-peer services may give positive ratings even when expectations are negatively disconfirmed. Study 4 will manipulate provider control to determine its effect on perceptions of integrity, while Study 5 will manipulate provider causality to determine its effect on perceptions of reliability. The studies will further test how social norms of gratitude and empathy in peer-to-peer services weaken the effect of reliability, but not integrity, on ratings when expectations are negatively disconfirmed.

8   Study 4

From the previous studies, expectancy disconfirmation is related to trust, which leads to ratings for peer providers. Trust assessments are comprised of perceptions of reliability (i.e., confidence that the provider has the ability to reliably deliver the required service

level) and integrity (i.e., confidence that the provider will honour their commitments fairly). However, in the case of *negative disconfirmation*, performance failure may not always lead to strong feelings about a provider's unreliability or lack of integrity. Study 4 explores the relationship between negative disconfirmation and integrity. It tests the proposition that negative disconfirmation leads to perceptions about a provider's lack of integrity only if the consumer believes that the provider could have controlled the negative outcome (P_8). If true, this could contribute to higher ratings in peer-to-peer services because some factors that would be controllable for commercial providers may be considered uncontrollable for peer providers. Peer providers are not professionals, and may not have the experience, ability, or financial resources to deliver expected service levels. If a consumer feels that a missed expectation is out of the provider's control, they may consider the provider to be unreliable, but not without integrity. The provider put forth their best effort in good faith (though nonetheless delivered unsatisfactory service).

Study 4 was also designed to test propositions P_{9a} and P_{9b}. I propose that social norms of gratitude and empathy contribute to ratings bias in peer-to-peer services because they motivate consumers to post higher ratings than may otherwise be deserved (i.e., P_{9a}). Gratitude and empathy may cause consumers to forgive providers for being unreliable but not if the provider lacks integrity (P_{9b}). If a provider places their own interest above the consumer's interest, it breaks the social exchange. Thus, consumers who determine that a provider lacks integrity should feel a desire to punish them with lower ratings. I manipulate provider control to determine whether this leads to lower perceptions of integrity and lower ratings. Table 11 lists the hypotheses to be tested in Study 4. Figure

11 shows the model to be tested in Study 4.

Table 11: Study 4 Hypotheses

|	
4.1	Negative disconfirmation is negatively related to perceptions of provider integrity
4.2	Negative disconfirmation is negatively related to perceptions of provider reliability
4.3	Provider control moderates the relationship between negative disconfirmation and integrity; this relationship will be stronger when provider control is higher
4.4	Social norms are positively related to ratings.
4.5	When expectations are negatively disconfirmed, social norms moderate the relationship between reliability and ratings; this relationship will be weaker when social norms are higher
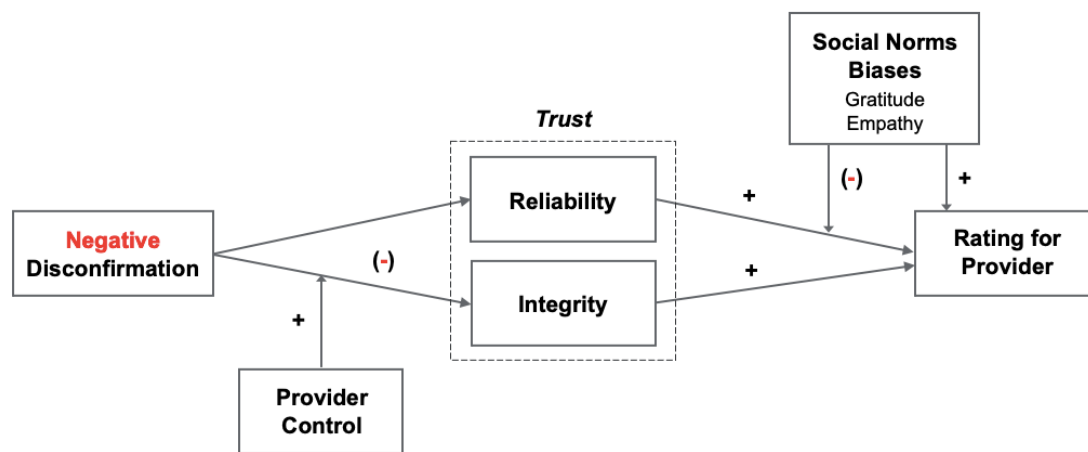
Figure 11: Model to be tested in Study 4

8.1 Method

8.1.1   Participants & Design

The experiment was conducted in the waiting area outside of a behavioral lab. It followed

a similar design to Study 1. While waiting to participate in an unrelated study, participants were intercepted by a confederate standing by a table with a tray of chocolate chip cookies. The confederate asked the participants if they would like to sample a chocolate chip cookie, and to take part in a survey about a new food service called Home Café that had recently been approved by the university. Two hundred and twenty-six undergraduate students agreed to participate in the survey. One participant was removed because they suspected that Home Café was not real and was part of a research study. The final sample included 225 participants.

Participants were randomly assigned to one of six conditions in a 2 (*Service Type*: commercial vs. peer-to-peer) x 3 (*Controllability*: controllable vs. uncontrollable vs. ambiguous) between-subjects design. The confederate described the Home Café service to the participants, and explained that the service was sampling some food items to get initial feedback and ratings for its website that would be launching to the public later that month. In the *commercial conditions*, Home Café was described as a new food service that prepared and sold home-style meals and snacks to busy students. The confederate described herself as a student employee of Home Café and was wearing a t-shirt with a Home Café logo. In the *peer-to-peer conditions*, Home Café was described as a new peer-to-peer food service that "works sort of like an Uber or an Airbnb for food. Individual students who like to cook can join the service and sell our home-style meals and snacks to other busy students". The confederate described herself as a student peer provider for the Home Café platform, and was wearing a plain t-shirt with no logo.

The confederate told participants that she would be sampling chocolate chip cookies, and that the confederate had personally baked the cookies as an employee of Home Café (or peer provider for Home Café). The confederate gestured toward a laptop computer on which appeared a webpage for the Home Café website (see Appendix M). The webpage was created using the Wix.com website builder. Although the webpage was fictional, it was designed to deceive participants into believing that it was a real website. A photo of the confederate appeared on the website, along with a picture of the chocolate chip cookies, a product name and product description, and an ingredient list. In the *commercial conditions*, participants were told that the website allows customers to see which Home Café staff member prepared their home-style meal or snack. In the *peer-to-peer conditions*, participants were told that the website allows customers to see the different meals offered by the peer providers, and that the product descriptions for each item were written by the individual peer provider.

Participants read the product description and then were given one chocolate chip cookie to sample. To elicit negative disconfirmation, the cookies were left out for a few days prior to the study so that they would be somewhat hard, dry, and stale when they were sampled. In the *controllable conditions*, the product name on the website was "Melt-in-your-mouth" cookies. The product description further promised that the cookies were yummy and moist, and that they were "ooey and gooey and melt-in-your-mouth". Because the cookies had been left to harden and stale, this description should not have matched the actual product experience. I expected that participants should infer that the description was misleading, and that this caused the missed expectations. Participants

should feel that the issue was controllable, especially in the peer-to-peer condition because the provider wrote the product description herself.

In the *uncontrollable conditions*, the product name and description was the same as in the controllable conditions, but the confederate mentioned that the university had told her that the research was being conducted earlier in the week. The confederate said that she baked the cookies for the original date, and so they may no longer be as fresh. In this case, the negative disconfirmation should be considered uncontrollable because it was caused by the university rather than the provider. In the *ambiguous conditions*, the product name was "Ultimate" cookies. The description further promised that the cookies were "sweet and satisfying, with chocolate chips in every bite". This product name and description did not reference the cookie texture. I expected that the staleness of the cookie should still lead to negative disconfirmation, but that it would not be clear whether the provider controlled the outcome. Perhaps the missed expectations were simply a matter of personal taste, or in the peer-to-peer conditions, perhaps the peer provider was simply not qualified or experienced enough to know that the cookies were too hard.

After sampling the cookie, participants were directed by the confederate to scan a QR code to activate an online survey from Home Café on their mobile phones. On the survey, participants were first asked to provide a *rating* for their experience on a five-star scale. They were told that this rating would be posted to the Home Café website. Participants next rated *expectancy disconfirmation* on a five-point scale. If a participant indicated that they experienced negative disconfirmation (i.e., 1 or 2 on the five-point scale), they were

then asked to indicate how strongly they felt that the negative experience was controllable by the provider. Specifically, *provider control* was a three-item measure ($\alpha = .92$): "My poor Home Café experience was definitely controllable by my provider," "My poor Home Café experience was definitely preventable by my provider," "My poor Home Café experience was definitely avoidable by my provider" (1 = Strongly Disagree, 5 = Strongly Agree). Finally, participants indicated their *satisfaction* (1 item), *trust* (four items; $\alpha = .87$), and *social norms* of gratitude and empathy (six items; $\alpha = .87$).

8.2 Results

I expected that perceived provider control would be significantly higher in the controllable condition than the other two conditions, and that it would be lowest in the uncontrollable condition. I further expected that provider control should affect the participant's feelings about the provider's integrity as per H$_{4.3,}$ such that integrity was lowest in the controllable condition. For reliability, I expected that it would be similarly low in the controllable and the ambiguous condition. In the ambiguous condition, although it is not clear whether the provider intended to deceive (because the product description did not describe the cookie dishonestly), the staleness of the cookie should nevertheless cause participants to feel that the provider is unreliable.

Unfortunately, the study failed to elicit sufficient levels of negative disconfirmation despite the fact that the cookie was left out to get stale. Only 5 participants (2.2%) indicated that the cookie was worse than expected. Four of these participants were in the

controllable condition, and one was in the uncontrollable condition. Thus it was not possible to assess the effectiveness of the controllability manipulation. It was also not possible to test proposition P_8 or P_{9b} because they specifically refer to the effects of *negative* disconfirmation. The specific hypotheses related to these propositions ($H_{4.1}$, $H_{4.2}$, $H_{4.3}$, and $H_{4.5}$) will instead be tested in Study 5.

The Summary Statistics and Correlation Matrix for Study 4 are in Appendix N. Descriptive statistics by condition are in Table 12. Comparisons between the variables of interest were tested with ANOVA. There were no significant differences between the peer-to-peer and commercial conditions for any of the measured variables. However, planned contrasts between the controllability conditions revealed some differences. Satisfaction was significantly lower in the controllable condition (M = 5.77) compared to the ambiguous condition (M = 6.21, t(203) = -2.07, $p < .05$) and marginally lower in the controllable condition compared to the two other conditions combined (t(203) = -1.91, $p < .10$). As expected, reliability did not significantly differ between the controllable condition (M = 4.47) and the ambiguous condition (M = 4.61, t(215) = -1.30, $p = .20$). I expected that reliability would be higher in the uncontrollable condition, but it did not differ between this condition (M = 4.54) and the other two conditions combined (t(215) = .04, $p = .97$). This is likely because the majority of participants (71%) indicated positive disconfirmation. Thus reliability was relatively high in all conditions. As expected, integrity was lower in the controllable condition compared to the other two conditions combined (t(215) = -1.77, $p < .10$), although the difference was marginally significant.

Table 12: Means and Standard Deviations for Study 4

	Commercial Controllable	Commercial Uncontrollable	Commercial Ambiguous	P2P Controllable	P2P Uncontrollable	P2P Ambiguous
	N = 33	N = 35	N = 36	N = 46	N = 39	N = 36
Rating	4.76 (.44)	4.74 (.51)	4.72 (.57)	4.70 (.55)	4.77 (.43)	4.78 (.42)
Expectancy Disc.	3.97 (.81)	3.91 (.85)	4.22 (.76)	3.89 (.95)	3.95 (.86)	4.03 (.77)
Provider Control	1.67			3.00 (1.52)		4.33
Satisfaction	5.93 (1.18)	6.11 (1.13)	6.28 (.96)	5.67 (1.57)	5.92 (1.40)	6.13 (.86)
Trust	4.27 (.88)	4.66 (.50)	4.61 (.56)	4.54 (.65)	4.45 (.49)	4.54 (.56)
Reliability	4.33 (.89)	4.67 (.51	4.63 (.61)	4.56 (.67)	4.42 (.64)	4.59 (.53)
Integrity	4.22 (.97)	4.64 (.54)	4.60 (.57)	4.51 (.72)	4.49 (.53)	4.50 (.54)
Social Norms	3.53 (.92)	3.98 (.80)	3.99 (.92)	3.79 (.98)	3.71 (.72)	3.82 (.79)

Next, I tested for differences in social norms of gratitude and empathy between conditions. I expected that social norms would be higher in peer-to-peer services, but results showed that they did not significantly differ between the peer-to-peer and commercial conditions ($M_{Commercial} = 3.85$, $M_{Peer} = 3.78$, $F = .36$, $p = .55$). Social norms also did not significantly differ between the three controllability conditions, but they were marginally lower ($t(207) = -1.96$, $p < .10$) in the controllable commercial condition compared to the other five conditions combined. Finally, a linear regression tested $H_{4.4}$. Supporting $H_{4.4}$, social norms were positively related to ratings ($\beta = .12$, $SE = .02$, $p < .01$).

8.3 Discussion

Disappointingly, this study did not elicit sufficient levels of negative disconfirmation.

Although the cookies were left out over several days to become stale, the majority of participants enjoyed their consumption experience. Because the levels of negative disconfirmation were so low, it was not possible to test most of the hypotheses. These hypotheses will be tested in Study 5. In retrospect, it may have been necessary to add some aversive tasting ingredients to the cookie to elicit negative disconfirmation. I did not want to do this because it was so important that participants believed that they were taking part in a real food service sampling experience. I was concerned that if the cookie's poor taste was too obvious, it would jeopardize the deception.

The only hypothesis that could be tested in Study 4 was $H_{4.4}$. This hypothesis was supported. Social norms were positively related to ratings. Thus, if social norms are higher in peer-to-peer services, they may contribute to the ratings bias. However, the results of this study showed that there was no significant difference in social norms for peer-to-peer compared to commercial conditions. In both the peer-to-peer and commercial conditions, the provider was described as a student, and it was made clear that the provider baked the cookies herself. Perhaps this level of personalization and personal connection engenders strong feelings of social norms even for employees of a commercial business. This may be another benefit of relationship marketing and personal selling that should be explored in future studies. Alternatively, there may be a difference between peer-to-peer services in which providers share their personal assets (homes, vehicles, tools, clothing etc.) and those in which they share their skills only. Consumers may have stronger feelings of gratitude and empathy when they are invited into a provider's personal home, and when providers entrust them with important personal

items. Those feelings may not be as strong for peer providers of skills-based services. This prediction should be explored in future studies. Finally, the relationship between social norms and ratings will be further tested in Study 5 to see if it still holds when a consumer experiences negative disconfirmation.
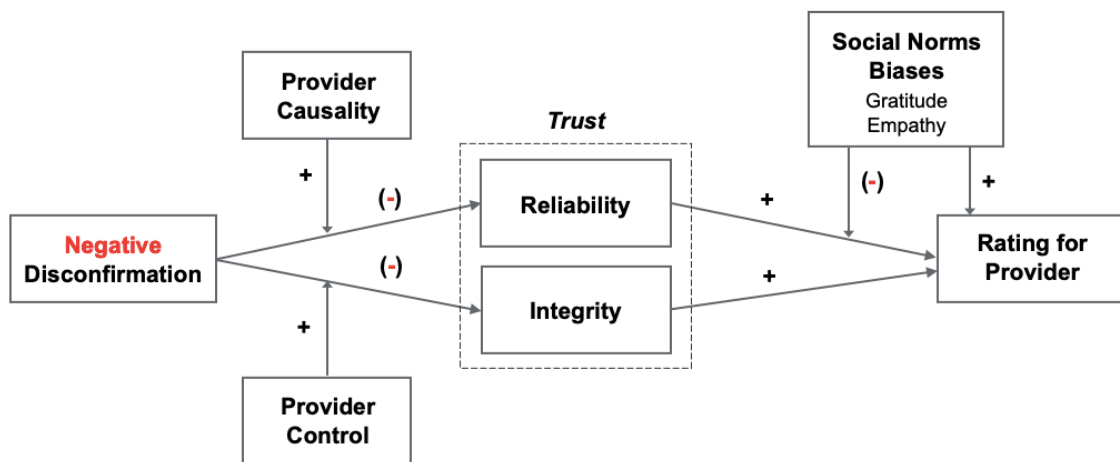
9   Study 5

Study 5 tests propositions P_7, P_8, P_{9a}, and P_{9b}. The main objective of Study 5 is to test the relationship between negative disconfirmation and reliability, and to explore how it is affected by clarity of standards and assessments of causal attribution (P_7). I predict that when standards of evaluation are less clear, as they are in peer-to-peer services, consumers will more likely blame themselves (rather than their provider) for causing missed expectations. This leads to higher ratings through higher perceptions of provider reliability. I manipulate clarity of standards to determine whether this helps attenuate the positive bias for peer-to-peer ratings.

Following from Study 4, this study also tests the proposition that negative disconfirmation leads to feelings that the provider lacks integrity only if they could have controlled the performance failure (P_8). Finally, I test how social norms of gratitude and empathy differentially affect the relationship of reliability and integrity with ratings (P_{9a} and P_{9b}.) Table 13 lists the hypotheses to be tested in Study 5. Figure 12 shows the model to be tested in Study 5.

Table 13: Study 5 Hypotheses

5.1	Clarity of standards is positively related to perceptions of provider causality
5.2	Negative disconfirmation is negatively related to perceptions of provider reliability
5.3	Negative disconfirmation is negatively related to perceptions of provider integrity
5.4	Provider causality moderates the relationship between negative disconfirmation and reliability; this relationship will be stronger when provider causality is higher
5.5	Provider control moderates the relationship between negative disconfirmation and integrity; this relationship will be stronger when provider control is higher
5.6	Social norms of gratitude and empathy are positively related to ratings.
5.7	When expectations are negatively disconfirmed, social norms moderate the relationship between reliability and ratings; this relationship will be weaker when social norms are higher

Figure 12: Model to be tested in Study 5

9.1 Method

Four hundred and one North American participants (211 women; $M_{age} = 32.09$ years) were recruited via the Prolific online research panel. Participants were randomly assigned

to one of four conditions in a 2 (*Service Type*: commercial vs. peer-to-peer) x 2

(*Standards*: undefined vs. defined) between-subjects design. Participants read that they

were planning a weekend trip to Seattle WA and needed to choose a mid-range

accommodation for a 2-night stay. Participants reviewed two different accommodation

listings on a fictional online booking site. Both of the options were listed for $109 per

night and both were described as a standard mid-range accommodation (i.e., "neither

basic nor premium"). The accommodation listings included a cover photo, a short

description of the unit, and a list of the amenities (see Appendix O). In the *commercial*

conditions, participants were told that the accommodations were listed on a website

called HotelEasy.com. The website names and logos were fabricated to control for the

fact that existing attitudes toward real brands could impact participant responses.

Participants were told that *HotelEasy.com* was an online booking aggregator similar to

Hotels.com or Booking.com, and the accommodations were offered from many different

(unnamed) hotels. In the *peer-to-peer conditions*, participants were told that the

accommodations were listed on a peer-to-peer home rental website called

HomeEasy.com. Participants were told that the accommodations on HomeEasy.com were

provided by individual homeowners, and that the platform was similar to Airbnb or

HomeAway. Further, in the commercial conditions, the accommodation providers were

listed as *managers*, and in the peer-to-peer conditions, the providers were listed as *hosts*.

Participants selected one of the two accommodation options. They were thanked for their

booking, and read about the accommodation standards on HotelEasy.com

(HomeEasy.com). In the *undefined standards conditions*, which served as the control

conditions, participants read that the company stood behind the quality of the accommodations listed on their website. In the *defined standards conditions* participants further read that the company asks all providers to adhere to the following standards: 1) accommodations should be extraordinarily well-kept, so that discerning guests can expect the rooms and amenities to be maintained in top condition; 2) accommodations should be thoroughly cleaned prior to rental, so that discerning guests can expect immaculate cleanliness; 3) accommodations should adhere to strict privacy, so that guests can relax without fear of undue noise or interruption. I expected that, compared to the undefined conditions, the standards of comparison in the defined standards conditions should be relatively more clear, and that this difference would be especially large for the peer-to-peer conditions because standards are relatively less clear in peer-to-peer services than commercial services. Participants clicked forward to accept that they read the standards, and then indicated their *expectations* (1 item), *perceived risk* (1 item), and *uncertainty* (two items; α = .89). These variables were measured the same as in prior studies. The correlation between perceived risk and uncertainty was moderate (r = .31, p < .01).

Next, participants read that they had arrived in Seattle and that they would be shown a series of photos that described their actual rental experience. Specifically, participants read that the photos represented what they saw when they walked around the rental accommodation. Participants were shown a total of 15 photos of their "experience" (see Appendix P). The photos advanced automatically after 5 seconds so that all participants spent the same amount of time viewing the images. The photos showed rooms and amenities that looked moderately well-kept, but with some wear and tear (i.e., scuffing on

the walls, small cracks on the ceiling etc.). There was a photo of a coffeemaker with an

"out of order sign". The rooms looked relatively but not immaculately clean. There was a

photo of a bathroom floor with a single hair on the ground. The collection of photos were

pretested to give an overall impression of a mid-range accommodation that was

moderately clean and well-kept, but not to the quality that one might expect.

After viewing their rental experience, participants were asked to provide an online rating

for the accommodation and its manager (host) on a five-star scale. Participants were told

that the rating would be posted to the company website. Next, participants rated

expectancy disconfirmation (1 item)*, satisfaction* (1 item)*, and trust* (four items; α = .94)*,

which were measured the same as in previous studies. Participants who indicated that

they experienced negative disconfirmation were then asked the *provider causality* (four

items; α = .73) and *provider control* (three items; α = .89) questions to determine

attribution.

All participants then rated the accommodation on several attributes based on how they

compared to what the participant would expect from a standard mid-tier accommodation.

The attributes were measured on seven-point semantic differential scales: *overall quality*

("Very LOW quality/Very HIGH quality"), *cleanliness* ("Very CLEAN/Very DIRTY"),

and *condition* ("Very POOR condition/Very GOOD condition"). Participants then

responded to *clarity of standards* measure which was rated on a seven-point semantic

differential scale. Specifically, participants were asked "How clear are you about the

standards of quality and service that should be expected from a manager (host) of a mid-

tier hotel (home) accommodation?" ("Not at all clear on the service standards/Completely clear on the service standards"). Finally, to measure *social norms* (6 items; α = .93)*,* participants were asked "if you were actually renting this accommodation, to what extent would you feel the following emotions?": grateful, appreciative, sympathetic, warm, compassionate, close (1 = Not at All, 5 = A Great Deal).

9.2 Results

9.2.1 Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 5 are in Appendix Q. Descriptive statistics by condition are in Table 14. As expected, expectations did not significantly differ between the peer-to-peer and commercial conditions (M_{Peer} = 5.62, $M_{Commercial}$ = 5.67, F(1, 399) = .34, *p* = .56). However, expectations were higher in the defined standards (M= 5.76) compared to the undefined standards conditions (M = 5.52, F(1, 399) = 7.54, *p* < .01). Planned contrasts revealed that this difference was significant in the commercial conditions ($M_{Defined}$ = 5.85, $M_{Undefined}$ = 5.48, t(397) = -3.00, *p* < .01) but not in the peer-to-peer conditions ($M_{Defined}$ = 5.67, $M_{Undefined}$ = 5.56, t(397) = -.89, *p* = .38). Thus, the changes I expected to see in the ratings between the peer-to-peer conditions were not affected by differences in expectations. Also as expected, perceived risk was significantly higher for the peer-to-peer conditions compared to commercial conditions (M_{Peer} = 3.32, $M_{Commercial}$ = 2.86, F(1, 399) = 13.47, *p* < .01), and uncertainty was marginally higher for the peer-to-peer conditions compared to commercial conditions (M_{Peer} = 2.64, $M_{Commercial}$ = 2.48, F(1, 399) = 2.95, *p* < .10).

Next, the variables of interest were compared between conditions. The stimuli successfully elicited negative disconfirmation without eliciting a large proportion of extremely negative opinions. Overall, 76.3% of participants experienced negative disconfirmation (i.e., 1, 2 or 3 on the seven-point scale) after viewing the accommodation photos. Of those who experienced negative disconfirmation, only 13.2% said that the experience was "much worse" than they expected. A total of 18% of participants experienced confirmed expectations, while 5.3% felt that the accommodation experience was better than they expected.

As expected, ratings were significantly higher for the peer-to-peer compared to the commercial conditions (M_{Peer} = 3.32, $M_{Commercial}$ = 2.86, $F(1, 399)$ = 13.47, $p < .01$). Planned contrasts revealed that, as expected, ratings were higher for the undefined peer-to-peer ($M = 3.24$) compared to the undefined hotel condition ($M = 2.90$, $t(397) = 2.72$, $p < .01$), and were not significantly different between the defined peer-to-peer ($M = 3.05$) compared to the defined hotel condition ($M = 2.91$, $t(397) = 1.12$, $p = .27$). This supports my prediction that implementing defined standards for peer-to-peer accommodations reduces positive bias compared to commercial accommodations. I expected that the mean rating in the undefined peer-to-peer condition would also be higher than in the defined peer-to-peer condition, but this difference was not significant ($t(397) = 1.52$, $p = .13$). Expectancy disconfirmation was significantly higher for the peer-to-peer compared to the commercial conditions (M_{Peer} = 2.93, $M_{Commercial}$ = 2.63, $F(1, 399)$ = 6.65, $p < .05$). This result was unexpected because there was no significant difference in expectations.

Contrast tests revealed that, compared to all of the other conditions, the undefined peer-to-peer condition had significantly higher disconfirmation ($t(397) = 2.25$, $p < .05$).

Table 14: Means and Standard Deviations for Study 5

	P2P Undefined	Commercial Undefined	P2P Defined	Commercial Defined
	N = 100	N = 100	N = 100	N = 101
Expectations	5.56 (.93)	5.48 (.82)	5.67 (.94)	5.85 (.82)
Perceived Risk	3.25 (1.34)	2.84 (1.21)	3.39 (1.26)	2.87 (1.26)
Uncertainty	2.66 (.96)	2.51 (1.00)	2.62 (.87)	2.45 (.90)
Rating	3.24 (.90)	2.90 (.86)	3.05 (.97)	2.91 (.80)
Expectancy Disc.	3.00 (1.16)	2.61 (1.11)	2.85 (1.24)	2.65 (1.03)
Provider Causality	4.92 (1.02)	5.06 (1.11)	4.93 (1.12)	5.38 (1.13)
Provider Control	5.81 (1.08)	5.36 (1.31)	5.81 (.85)	5.34 (1.25)
Satisfaction	3.96 (1.63)	3.37 (1.50)	3.63 (1.56)	3.22 (1.38)
Trust	3.92 (1.29)	3.60 (1.21)	3.97 (1.27)	3.63 (1.18)
Reliability	3.96 (1.32)	3.60 (1.29)	4.01 (1.34)	3.59 (1.23)
Integrity	3.88 (1.34)	3.61 (1.25)	3.92 (1.26)	3.66 (1.21)
Quality	3.84 (1.38)	3.22 (1.25)	3.56 (1.32)	3.15 (1.16)
Cleanliness	3.75 (1.60)	3.23 (1.36)	3.49 (1.42)	3.33 (1.33)
Condition	3.87 (1.45)	3.06 (1.24)	3.69 (1.35)	3.14 (1.21)
Social Norms	2.10 (.89)	1.82 (.78)	2.26 (.99)	1.84 (.79)
Clarity of Standards	4.63 (1.47)	4.90 (1.41)	5.01 (1.34)	4.98 (1.54)

Next, I compared social norms, clarity of standards, and provider causality between conditions. I expected that social norms of gratitude and empathy would be higher in peer-to-peer services. As expected, social norms were significantly higher in the peer-to-

peer compared to the commercial conditions (M_{Peer} = 2.18, $M_{Commercial}$ = 1.83, F(1, 398) = 16.86, p < .01). Next, I expected that clarity of standards would be lower in the peer-to-peer compared to commercial conditions, and higher in the defined compared to undefined standards conditions. I further expected that the differences between the defined and undefined conditions would be significant only for peer-to-peer services because standards should be relatively clear for commercial accommodations even if they are undefined. Clarity of standards did not significantly differ between the peer-to-peer (M = 4.82) and commercial conditions (M = 4.94, F(1, 398) = .70, p = .40) or between the defined and undefined standards conditions ($M_{Defined}$= 5.00, $M_{Undefined}$ = 4.76, F(1, 398) = 2.53, p = .11). However, planned contrasts revealed that, as expected, standards were more clear in the defined peer-to-peer conditions (M = 5.01) compared to the undefined peer-to-peer conditions (M = 4.63, t(396) = -1.86, p < .10; i.e., marginal significance) while they did not differ between the defined hotel conditions (M = 4.98) and the undefined hotel conditions (M = 4.90, t(396) = .15, p = .88). Providing some support for my predictions, the standards of evaluation were significantly less clear in the undefined peer-to-peer condition compared to all other conditions combined (t(396) = -2.00, p < .05).

Next, I expected that provider causality would be lower in the peer-to-peer compared to the commercial conditions, and higher in the defined compared to undefined standards conditions because of the expected differences in the clarity of standards. As expected, provider causality was significantly lower in the peer-to-peer (M = 4.92) compared to the commercial conditions (M = 5.22, F(1, 398) = 5.56, p < .05). However, provider causality

did not significantly differ between the defined and undefined standards conditions

($M_{Defined}$= 5.16, $M_{Undefined}$ = 5.00, F(1, 398) = 1.75, *p* = .19).

9.2.2   Tests of Hypotheses

Next, the hypotheses were tested. Hypothesis $H_{5.1}$ was tested with linear regression.

Supporting $H_{5.1}$, there was a significant positive relationship between clarity of standards

and provider causality (β = .16, SE = .04, p < .01) such that provider causality is

considered to be stronger when standards of evaluation are more clear. The remaining

hypotheses were tested using only the data from participants who experienced negative

disconfirmation (N = 305). With linear regression, I first tested the relationship between

negative disconfirmation and the two dimensions of trust: reliability and integrity.

Supporting $H_{5.2}$ there was a significant negative effect of negative disconfirmation on

reliability (β = -.73, SE = .08, *p* < .01) such that as negative disconfirmation becomes

stronger (i.e., as expectancy disconfirmation becomes lower, from 3 to 2 to 1), perceived

provider reliability is lower. Supporting $H_{5.3}$ there was a significant negative effect of

negative disconfirmation on integrity (β = -.72, SE = .08, *p* < .01). As negative

disconfirmation becomes lower, perceived provider integrity is lower.

Hypothesis $H_{5.4}$ predicted that provider causality strengthens the effect of negative

disconfirmation on reliability. I conducted a moderation analysis with negative

disconfirmation as the predictor, reliability as the dependent variable, and provider

causality as the moderator (PROCESS Model 1; Hayes 2018). Supporting $H_{5.2}$, there was

a significant negative relationship between negative disconfirmation and reliability (β = -.62, SE = .07, $p < .01$) such that as disconfirmation becomes more negative, reliability is similarly lower. Further, there was a significant negative main effect of provider causality on reliability, such that if consumers believe that providers have caused performance failure, then reliability will be lower (β = -.28, SE = .05, $p < .01$). However, the predicted interaction was not significant (β = .00, SE = .06, $p = .98$). Therefore $H_{5.4}$ was not supported, but the overall premise was confirmed; when performance is below expectations, a stronger feeling of provider causality leads to lower perceptions of provider reliability.

Hypothesis $H_{5.5}$ predicted that provider control strengthens the effect of negative disconfirmation on integrity. From a moderation analysis, there was a significant negative relationship between negative disconfirmation and integrity (which supports $H_{5.3}$; β = -.39, SE = .08, $p < .01$) such that as disconfirmation becomes more negative, integrity is lower. Additionally, there was a significant negative main effect of provider control on integrity, such that if consumers believe that providers have more control over the performance failure, then integrity will be lower (β = -.37, SE = .05, $p < .01$). Further, supporting $H_{5.5}$, there was a significant interaction (β = -.29, SE = .08, $p < .01$). The more that consumers feel that their provider had control over a performance failure, the stronger is the negative effect from negative disconfirmation on integrity (see Figure 13). An analysis of conditional effects revealed that the negative effect of negative disconfirmation on integrity is significant if provider control is *higher* (M_{+1SD} = 6.80; β = -.74, SE = .12, $p < .01$) or *medium* (M_{Mean} = 6.00; β = -.51, SE = .09, $p < .01$). Negative

disconfirmation does not lead to lower perceptions of integrity if consumers feel that the providers had *lower* control over the performance failure (M_{-1SD} = 4.67; β = -.13, SE = .12, p = .25). Further, a Johnson-Neyman analysis showed that the negative effect of negative disconfirmation on integrity was significant only when provider control was at a value of 4.91 (out of 7) or higher (approximately 79% of participants). Conversely, if participants believe that provider control is relatively low, then negative disconfirmation does not affect their perceptions of the provider's integrity.
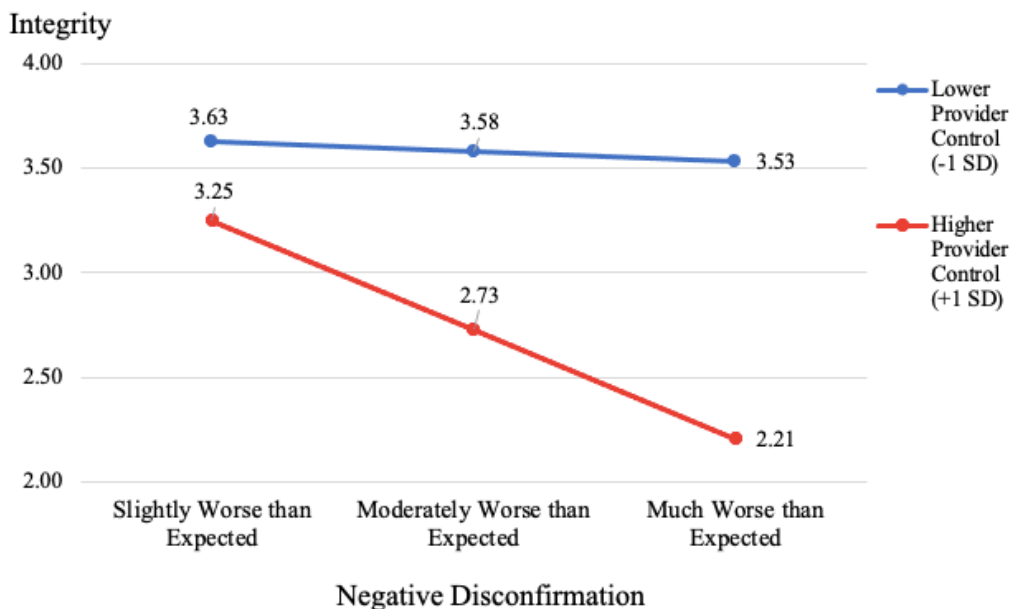
Figure 13: The Effect of Negative Disconfirmation on Integrity at Different Levels of Perceived Provider Control

.

Finally, $H_{5.6}$ and $H_{5.7}$ were tested. I expected that social norms of gratitude and empathy would add to ratings bias by motivating consumers to give higher ratings to their providers. Further, I expected that consumers would forgive providers for unreliable service if social norms were high. That is, social norms should moderate the relationship

between negative disconfirmation and unreliability, such that the effect of unreliability on ratings is weaker. Social norms should *not* affect the relationship between integrity and ratings, because consumers should feel justified giving a low rating to a provider if they feel that they lacked integrity. From a moderation analysis (see Figure 14) there was a significant positive main effect of reliability on ratings (β =.18, SE = .04, p < .01). Supporting $H_{5.6}$, there was a significant positive main effect of social norms on ratings (β = .46, SE = .06, p < .01). Finally, supporting $H_{5.7}$, there was a significant negative interaction (β = -.10, SE = .04, p < .05, d = .28). When expectations are negatively disconfirmed, the effect of reliability on ratings is weaker when social norms are higher. Further, a Johnson-Neyman analysis showed that (when expectations are negatively disconfirmed), reliability affects ratings only when social norms are at a value of 2.59 (out of 7) or lower (approximately 84% of participants).To show that social norms affected only reliability and not integrity, a moderation analysis was conducted with integrity as the predictor in place of reliability. As expected, the interaction between integrity and social norms on ratings was not significant (β = .06, SE = .04, p = .18). Thus social norms do not weaken the effect of a lack of integrity on ratings.
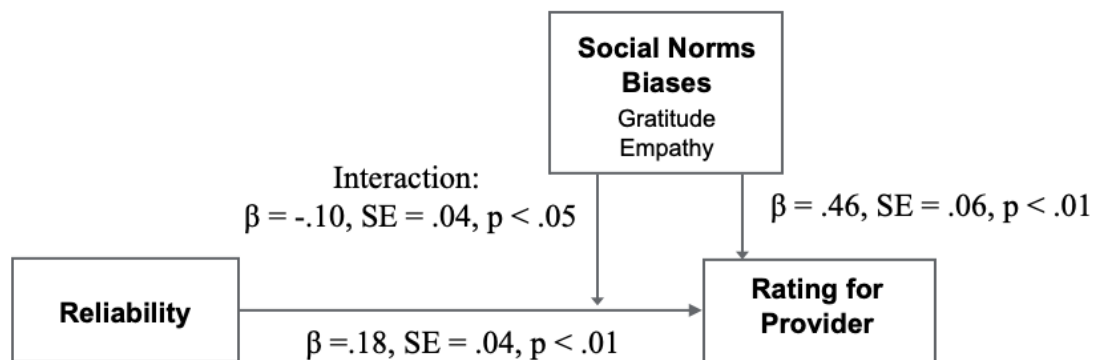

Figure 14: Results of Study 5 Moderation Analysis

9.3 Discussion

Study 5 successfully elicited sufficient levels of negative disconfirmation from participants. This allowed me to test all of the hypotheses. As expected, negative disconfirmation leads to feelings that providers are unreliable. Providers are considered to be especially unreliable if they are believed to have caused a consumer's missed expectations. Negative disconfirmation also leads to feelings that a provider lacks integrity, but not if consumers feel that providers had relatively low control over the performance failure. Further, as expected, gratitude and empathy cause consumers to forgive providers for unreliable service. This leads to higher ratings even when expectations are negatively disconfirmed, which contributes to the positive ratings bias. However, gratitude and empathy do not affect the relationship between integrity and ratings. If a consumer feels that a provider lacks integrity, it should lead to low ratings because providers have shown that they place their own interests above those of the consumer. Leveraging these findings, Study 6 will test a novel ratings system to determine whether it will help to attenuate the positive ratings bias by increasing perceived provider control.

Study 5 also demonstrated that evaluation standards are less clear in peer-to-peer services, leading to lower perceptions of provider causality for performance failures in peer-to-peer services compared to commercial services. Peer-to-peer platforms can attenuate the positive ratings bias by helping consumers to better understand how to evaluate peer-to-peer services. This can be done by introducing defined standards for peer-to-peer services. Importantly, these defined standards did not raise the level of

expectations for peer-to-peer consumers, but rather, they simply made the standards of evaluation more clear. Study 5 demonstrated that when standards of evaluation in peer-to-peer services are clearly defined, the difference in ratings between peer-to-peer and commercial services is removed.

10   Study 6

Studies 1 to 5 demonstrated that consumers of peer-to-peer services evaluate service experiences based in part on whether they trust their provider. Trustworthy providers honour their commitments and meet consumers' expectations. If expectations are met, even if the experience is not completely satisfying, consumers may give providers a high rating, signaling trustworthiness. When expectations are disconfirmed, consumers may only feel able to justify giving a low rating if they believe that a provider caused the missed expectations *and* also had control over the outcome. In this case, the provider could have prevented the issue, but did not; thus it can be determined that the provider lacks integrity. A lack of integrity negates the positive effects of gratitude and empathy, and should lead to low ratings for peer providers.

Leveraging these learnings, the objective of Study 6 is to test a novel ratings system that would attenuate the positivity bias by increasing a provider's perceived controllability. In many peer-to-peer platforms, consumers are asked to rate providers on a variety of attributes. For example, Airbnb asks consumers to rate providers on communication, cleanliness, location, check-in, and value. In my proposed system, providers would rate

themselves on these or other attributes. The provider-authored ratings act as an implicit service commitment by providers for each attribute, and consumers would draw on those ratings to help make purchase decisions. For example, an Airbnb consumer in the new system may decide to purchase from a provider after learning that the provider rated themselves and their property at 5-stars for cleanliness, 5-stars for communication and 4-stars for location.

Provider-authored ratings reduce information asymmetry, because providers have complete information about their quality. Provider-authored ratings thus give consumers a clear standard against which they can set their expectations. This is because providers have the ability and are expected to rate their attributes fairly and accurately. If actual performance is lower than the provider-authored ratings, the issue should be assumed to be controllable by the provider. Consumers should feel that the provider intended to deceive by giving artificially high ratings that were not commensurate with their actual performance level.

If a provider explicitly communicates that a consumer can expect a certain level of cleanliness (for example) but does not deliver it, it implies intentionality. This may be especially true for providers that rate themselves at five-stars, because five-star ratings mean that the service level should be perfect, or could not be higher. It is *not* difficult to evaluate whether performance was or was not perfect. If a provider fails to match this performance expectation, the consumer should feel that the provider was not honest in their assessment. The provider lacks integrity because they did not perform a fair

transaction. When consumers feel that they were treated unfairly, and that the provider lacks integrity, they should feel justified to rate the provider poorly. In turn, this should help to attenuate the positive ratings bias. On the other hand, if a provider rates themselves at four-stars for a particular attribute, but delivers something less, consumers may feel that the provider was relatively honest. Thus the effect of provider-authored ratings should be reduced. Table 15 presents the hypotheses to be tested in Study 6.

Table 15: Study 6 Hypotheses

6.1	Compared to consumer-authored ratings, provider-authored ratings lead to higher perceived provider control
6.2	The rating score moderates the relationship between provider-authored ratings and provider control; this relationship is stronger when the rating score is five-stars
6.3	When the rating score is five-stars, provider-authored ratings lead to lower perceptions of provider integrity compared to when ratings are consumer-authored
6.4	When the rating score is five-stars, provider-authored ratings lead to lower post-experience ratings compared to when pre-experience ratings are consumer-authored

10.1 Method

10.1.1 Participants & Design

Two hundred undergraduate students (120 women; M_{age} = 18.22) were recruited for course credit (80 women; M_{age} = 19.85 years). Participants were randomly assigned to one of four conditions in a 2 (*Rating Score*: Five-star vs. Mixed) x 2 (*Rating Author*: Consumer vs. Provider) between-subjects design which included a control condition. The

study was completed in two phases and followed a similar design to studies 2A and 2B.

In the first phase, participants received an email from the lab manager with the study

instructions. The email stated that the university was considering hiring a graphic design

company ("Netwrk") to provide design services for students, and that the purpose of the

study was to test the company's services for logo design. Participants were provided with

a link to the first of two surveys.

After clicking to begin the survey, participants were asked to review the digital profiles of

three freelance graphic designers who were using the Netwrk platform (see Appendix R),

and to choose one of those designers to work on their logo project. Each profile contained

a photo of the designer, a short description about their creative interests and design

experience, and a set of ratings for four different attributes: 1) originality, 2) level of

customization, 3) attention to detail, and 4) speed. Each of the attributes had been rated

on a five-star scale. In the *five-star ratings conditions* each of the four attributes (i.e.,

originality, level of customization, attention to detail, and speed) were rated at 5-stars

(i.e., a perfect rating) for all three of the graphic designers. In the *mixed ratings

conditions*, three of four attributes were rated at 4-stars and the fourth was rated at 5-stars.

I expected that, compared to the five-star ratings conditions, the mixed ratings would set

relatively lower quality expectations, and might be considered to be more accurate and

honest.

In the *consumer-authored ratings condition*s, the text "Consumers rate my services as"

was placed above the attribute ratings on the designers' profiles. Participants were told

that the attribute ratings represented an aggregate score of the ratings provided by

previous consumers who had worked with each particular designer. Participants were

reminded that ratings from previous consumers help new consumers understand the

quality and service that they can expect from the designer. In the *provider-authored*

*ratings condition*s, the text "I rate my services as" was placed above the attribute ratings.

Participants were told that the ratings for the attributes were given by the provider

themselves, to help consumers understand the quality and service that they can expect

from the designer. I expected that, compared to the consumer-authored ratings conditions,

participants who experienced negative disconfirmation in the provider-authored ratings

conditions would assess provider control higher. This is because the provider set the

participant's expectations by giving the ratings himself. I also included a *control*

condition in which the online listings for the designers did not include any ratings. This

was to control for the anchoring effect of prior ratings. It also allowed me to determine a

more accurate evaluation of designer performance which I could then compare to the

final evaluations in the other conditions to calculate an estimate of ratings bias.

Participants chose one of the three designers. I created a dummy variable to identify

which designer was chosen by each participant (0 = No, 1 = Yes). These variables were

used as covariates in the analyses. Next, participants were asked about their feelings

about the upcoming experience including their *expectations* (1 item), *perceived risk* (1

item), and *uncertainty* (two items; $\alpha = .71$), which were measured the same as in previous

studies. The correlation between perceived risk and uncertainty was moderate ($r = .40$, p

< .01). These measures were included to confirm that the different conditions did not

significantly affect the participants' feelings about the provider prior to the experience, and to ensure that any differences in expectancy disconfirmation and ratings were not due to differences in expectations, risk or uncertainty. Finally, participants provided their email address and were told that their chosen Netwrk designer would contact them via email later that day to begin the design process.

The second phase of the study began approximately one hour after the participant had completed the first survey. Posing as the chosen Netwrk designer, I sent an email to the participant to start the design process. The designer introduced himself and explained that he needed some information from the participant: their first and last name along with a key phrase or motto to include in the design; some direction on the style of logo that they wanted including whether they preferred it to be colourful or neutral, and clean/simple or detailed/artistic; and to highlight their interests from a list of choices including sports, music, art and business, which would help the designer determine a theme for the logo. Participants were also asked to provide additional background information that could be used to customize the design. The answers to the style and interest questions were used to select one of six logo design templates that would be used for the new design. I created a dummy variable to identify which template was given to each participant (0 = No, 1 = Yes). These variables were used as covariates in the analyses.

Participants responded to the email and provided the requested information. They received a thank you reply from the designer and were told that the designer would deliver their new design within 24 hrs, per the Netwrk service policy. The following day,

participants received a third email from the designer with the new logo. As in Studies 2A

and 2B, the email was delivered approximately one hour after the end of the 24-hour

window, thus failing to meet the service commitment. The designer apologized, and

stated that he was delayed because he was working on another project. I expected that

this would help to elicit negative disconfirmation, and specifically, that it would impact

the evaluation of the "speed" attribute. To further elicit negative disconfirmation I

misspelled one word in each participant's life motto (see Appendix S for examples) and

expected that this would impact the participants' assessment of the "attention to detail"

attribute. Finally, to get a range of evaluations for the "originality" and "level of

customization" attributes, I ignored the additional background information that

participants provided, and did not use it to customize the logo in any way.

The email also included a link to the second survey, on which participants could evaluate

the service experience. First, participants were asked to provide a *rating* on a 5-star scale.

They were told that the rating would be assigned to the designer, and would be shared

with Netwrk so that they could post it on their website. Next, participants were reminded

of the ratings scores that previous consumers (or the provider himself) gave the designer

for each of the four attributes. With these in mind, participants were asked to rate the

designer on a 5-star scale for each of *originality*, *customization*, *detail,* and *speed*. I

aggregated the scores on these four items to create a new variable called *attribute ratings*.

Participants next indicated *expectancy disconfirmation* (1 item) which was measured the

same as in previous studies. For participants who indicated that they experienced negative

disconfirmation, *provider control* was measured the same as in previous studies (three

items; α = .90.). Finally, *satisfaction* (1 item)*,* and *trust* (including the four items representing the dimensions of reliability and integrity; α = .91) were measured the same as in previous studies.

10.2 Results

10.2.1 Differences Between Conditions

The Summary Statistics and Correlation Matrix for Study 6 are in Appendix T. Descriptive statistics by condition are in Table 16. The variables of interest were compared between conditions with ANCOVA. As expected, the manipulations of the factors had no significant effect on expectations, perceived risk, and uncertainty. These variables did not significantly differ between any of the conditions. Disconfirmation was significantly higher in the five-star ratings conditions (M = 4.85) compared to the mixed ratings conditions (M = 4.24, $F(1, 151) = 7.19$, $p < .01$) and also in the five-star ratings conditions compared to the control condition (M = 3.73, $F(1, 111) = 14.63$, $p < .01$). This suggests that participants' evaluations after the experience may have been positively influenced by the high ratings that they saw before their experience. Relatedly, satisfaction was marginally higher in the five-star ratings conditions (M = 5.44) compared to the mixed ratings conditions (M = 5.03, $F(1, 151) = 3.41$, $p < .10$) and was significantly higher in the five-star ratings conditions compared to the control condition (M = 4.40, $F(1, 111) = 15.75$, $p < .01$). Disconfirmation did not differ between the consumer-authored (M = 4.65) and provider-authored conditions (M = 4.44, $F(1, 151) = .77$, $p = .38$), and satisfaction also did not differ between the consumer-authored (M =

5.30) and provider-authored conditions (M = 5.16, $F(1, 151) = .40$, $p = .53$)

Table 16: Means and Standard Deviations for Study 6

	Five-Star Rating Consumer N = 40	Five-Star Rating Provider N = 40	Mixed Rating Consumer N = 40	Mixed Rating Provider N = 40	No Rating (Control) N = 40
Expectations	5.50 (.93)	5.25 (1.10)	5.20 (.99)	5.30 (.94)	5.40 (.87)
Perceived Risk	2.75 (1.30)	2.93 (1.35)	2.63 (1.17)	3.03 (1.49)	2.85 (1.41)
Uncertainty	2.51 (.92)	2.64 (.88)	2.71 (.94)	2.94 1.16)	2.85 1.01)
Rating	4.58 (.71)	4.22 (.97)	4.18 (.75)	4.13 (.88)	3.55 (1.09)
Originality	4.63 (.59)	4.58 (.68)	4.22 (.73)	4.08 (.80)	3.45 (.99)
Customization	4.40 (.90)	4.08 (.97)	3.93 (.97)	3.88 (.99)	3.18 (1.08)
Detail	4.30 (.94)	3.73 (1.38)	3.65 (1.12)	3.70 (1.09)	3.00 (1.11)
Speed	4.63 (.74)	4.45 (.82)	4.18 (.84)	4.35 (.66)	3.93 (1.10)
Attribute Ratings	4.49 (.66)	4.21 (.71)	3.99 (.73)	4.00 (.64)	3.39 (.82)
Expectancy Disc.	5.05 (1.48)	4.63 (1.62)	4.25 (1.45)	4.25 (1.41)	3.73 (1.54)
Provider Control	5.35 (.50)	6.00 (.86)	4.67 (.81)	4.00 (2.24)	5.47 (.89)
Satisfaction	5.58 (1.45)	5.30 1.45)	5.05 (1.43)	5.00 1.47)	4.40 (1.39)
Trust	6.14 (.90)	5.88 1.01)	5.68 (.92)	5.77 (1.15)	5.22 (1.19)
Reliability	6.11 (1.05)	5.98 (1.11)	5.74 (.91)	5.74 (1.15)	5.14 (1.37)
Integrity	6.18 (.81)	5.79 (1.02)	5.63 (1.05)	5.80 (1.18)	5.24 (1.12)

10.2.2 Tests of Hypotheses

Next, the hypotheses were tested. Hypothesis $H_{6.1}$ was tested with ANCOVA. I expected that perceptions of provider control would be higher in the provider-authored ratings

conditions because the providers set the expectations themselves based on the ratings they gave. Providing some support for $H_{6.1}$, provider control was marginally higher in the provider-authored (M = 5.52) compared to the consumer-authored ratings conditions (M = 4.19; $F(1, 10) = .35$, $p < .10$). The low number of observations may have impacted this result. Despite the fact that the text in each logo was misspelled, and that the logos were delivered late, only 41 participants in the non-control conditions (26%) experienced negative disconfirmation. Future studies should test this hypothesis with a larger sample.

Next, a moderation analysis (PROCESS Model 1, Hayes 2018) was used to test $H_{6.2}$. I expected that perceptions of provider control would be especially high when the provider-authored ratings were five-stars (compared to mixed ratings) because the provider in the mixed ratings conditions may be perceived to be more honest about their relative lack of abilities and thus less responsible for the poor performance. The main effect of provider-authored ratings was not significant ($\beta = -.16$, SE = .68, $p = .82$) but the main effect of the five-star rating score was marginally significant ($\beta = 1.4$, SE = .70, $p < .10$). The predicted interaction between rating author and rating score was not significant ($\beta = 1.32$, SE = 1.42, $p = .37$) which again likely was impacted by the low number of observations. However, a planned contrast revealed that provider control was significantly higher in the provider-authored five-star ratings condition compared to the three other (non-control) conditions combined ($F(1,10) = 7.33$, $p < .05$; See Figure 15). This provides some support for the prediction from hypothesis $H_{6.2}$.
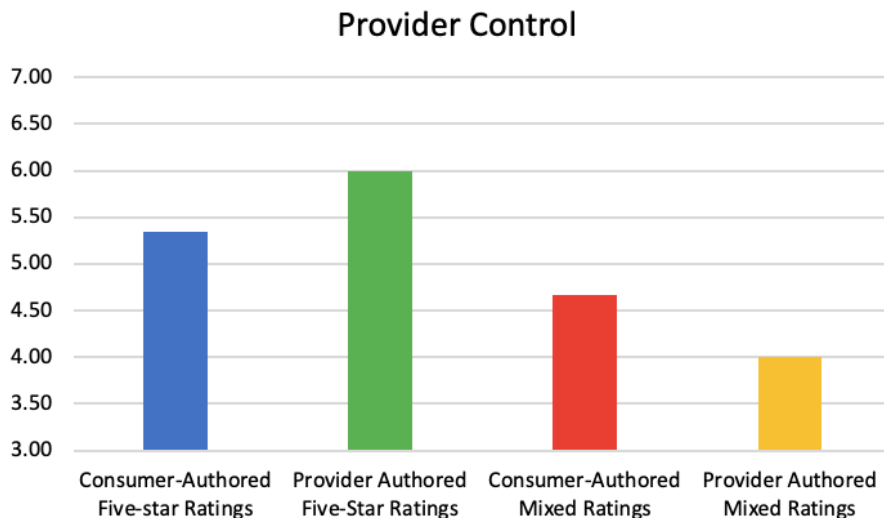
Figure 15: Comparison of Provider Control Across Conditions

Next, the final two hypotheses were tested. I expected that the higher level of perceived provider control for missed expectations in the provider-authored five-star ratings condition would lead to lower perceived integrity ($H_{6.3}$) and subsequently lower ratings ($H_{6.4}$) compared to the consumer-authored five-star condition. The hypotheses were tested with linear regression for the five-star ratings conditions. The chosen provider and the logo design template were used as covariates. Supporting $H_{6.3}$, integrity was significantly lower when the five-star ratings were provider-authored compared to when they were consumer-authored ($\beta = .-41$, $SE = .20$, $p < .05$; see Figure 16). As expected, subsequent analyses revealed that reliability ($\beta = .-15$, $SE = .23$, $p = .53$) and satisfaction ($\beta = .-27$, $SE = .31$, $p = .39$) did not differ between the provider-authored and consumer-authored ratings conditions. Supporting $H_{6.4}$, the ratings (i.e., the final ratings that were chosen by the participants after their experience) were significantly lower when the pre-experience five-star ratings were provider-authored compared to when they were consumer-authored

(β = .-37, SE = .18, *p* < .05). This difference was also significant for the assessment of

attention to detail (β = .-58, SE = .24, *p* < .05) and was marginally significant for

customization (β = .-34, SE = .20, *p* < .10), and aggregate ratings (β = .-29, SE = .15, *p* <

.10), such that these ratings were lower in the provider-authored conditions. There was no

significant difference for the assessments of originality (β = .-06, SE = .14, *p* = .66) or

speed (β = .-19, SE = .18, *p* = .29). Disconfirmation was also not significantly different

between these two conditions (β = .-40, SE = .34, *p* = .24).
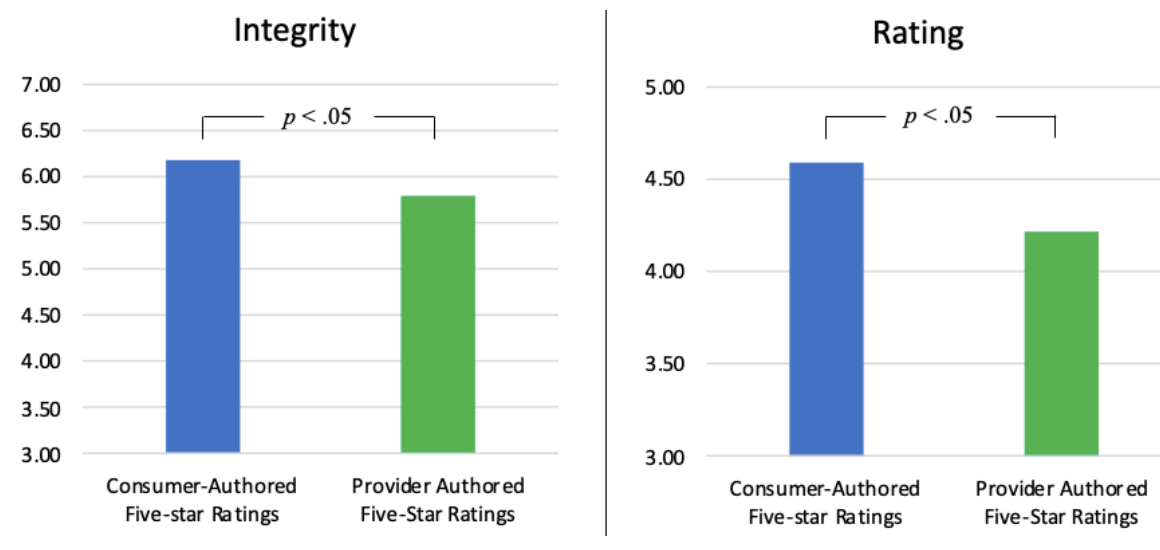
Figure 16: Comparison of Consumer-Authored and Provider-Authored Five-Star Ratings
Conditions for Integrity and Rating

Finally, ANCOVA was used to probe the differences in ratings between conditions, and

to estimate the size of the positivity bias. Ratings were significantly higher in the five-star

conditions (M = 4.41) compared to the mixed ratings conditions (M = 4.14, F(1, 151) =

4.25, *p* < .05). Aggregate ratings were also significantly higher in the five-star conditions

(M = 4.35) compared to the mixed ratings conditions (M = 3.99, F(1, 151) = 10.97, *p* <

.01). This supports prior research that has shown that ratings behavior is significantly influenced by prior posted ratings (e.g., Moe & Trusov, 2011; Schlosser, 2005).

Ratings were significantly lower in the control condition (M = 3.55) compared to all other conditions combined (M = 4.28, F(1, 191) = 21.74, p < .01). Planned contrasts revealed that the rating in the control condition was significantly lower than in each of the other conditions individually. As expected, on an average basis, the size of the ratings difference is largest in the consumer-authored five-star ratings condition (M = 4.58) compared to the control condition (M = 3.55), at over one point (i.e., 1.03) on the five-star scale. Most reviews that consumers see when they are selecting providers in real peer-to-peer platforms are consumer-authored five-star ratings, so this difference of one point provides a good starting point for an estimate of the size of the positive ratings bias. Aggregate ratings were also significantly lower (M = 3.39) in the control condition compared to all other conditions combined (M = 4.17, F(1, 151) = 37.90, p < .01). I expected that the ratings bias in aggregate ratings would be lower than in actual ratings, but this was not the case. Again the difference was greatest between the consumer-authored five-star ratings condition (M = 4.49) compared to the control condition (M = 3.39), at over one point (i.e., 1.10) on the five-star scale.

10.3 Discussion

Study 6 tested the effects of a novel ratings system on attenuating the positive bias that was based on the learnings from Study 5. Peer-to-peer consumers will forgive unreliable

providers but not will not forgive providers who lack integrity. The novel system was designed to elicit negative feelings of integrity after negative disconfirmation, by increasing the perception of provider control over performance failures. Study 6 did not elicit as much negative disconfirmation as anticipated, but nevertheless the study supports the predictions. Providers are perceived to have higher control over missed expectations when the pre-experience ratings are provider-authored and five-stars. This leads to lower integrity and lower ratings for the provider-authored ratings compared to the consumer-authored ratings in the five-star condition. Thus, peer-to-peer platforms can reduce the positivity bias by assigning responsibility to providers to set their own attribute ratings.

The study further attempted to calculate the size of the positivity bias. Results showed that ratings in the control condition were between 0.5 points to 1 point lower than the other conditions (on a five-star scale). This difference is likely a conservative estimate of the positivity bias, because a portion of participants may not have felt sure that the ratings exercise was real, and others may not have felt a strong need to justify ratings because they are not members of the peer-to-peer platform and providers cannot harm them with low ratings as retribution.

Chapter 3

11   General Discussion

The goal of this thesis was to explain the important problem of positive bias in peer-to-peer ratings, and to provide potential solutions for platforms. To tackle this problem, I endeavored to develop a deep understanding of how consumers evaluate peer-to-peer experiences, and how unique contextual factors in peer-to-peer exchanges (compared to commercial exchanges) affect the relationship between performance evaluation and ratings. In this chapter, I will explicate the main contributions of the thesis as a whole, explore the limitations and directions for future research, and provide final thoughts about the positive bias in peer-to-peer ratings. For a summary of results of the hypothesis tests in each study, see Appendix U.

11.1 Contributions

In this thesis, I have developed empirical support for the notion that trust is an important factor for peer-to-peer evaluations, and that evaluations of trust, in the form of reliability and integrity, contribute to the positive bias in peer-to-peer ratings. The thesis makes three main contributions. The first contribution is the demonstration that expectancy disconfirmation leads to evaluations of trust (in addition to satisfaction). Research on product and service evaluation is most often informed by the well-established expectancy disconfirmation process (Oliver, 1980, 2010). Consumers compare a provider's

performance against their prior expectations, and the resultant satisfaction or dissatisfaction leads to a variety of behaviors including word-of-mouth (Anderson, 1998; Homburg et al., 2005). The model has been supported across many contexts. However, I demonstrate that the model works differently for peer-to-peer services. Study 1 shows that, in a peer-to-peer service, the consumer's determination of whether the peer provider met or did not meet their expectations has an effect on provider ratings that is above and beyond the effect of satisfaction. This suggests that there may be missing mediators between expectancy disconfirmation and ratings. I demonstrate that one of these mediators is trust.

Trust is closely related to satisfaction, and is also associated with word-of-mouth (Ranaweera & Prabhu, 2003). However, the link between expectancy disconfirmation and trust is not well-established in extant research. I explain why expectancy disconfirmation leads to trust, and demonstrate the moderators of this effect (uncertainty, provider causality, provider control). Study 2B and Study 3 show that when uncertainty is high, a provider can demonstrate their trustworthiness by meeting the consumer's expectations. Under conditions of uncertainty, consumers must place their trust in the provider and in their claims, with hope that they will not act with opportunism. If the consumer's actions are rewarded, and the provider meets their promised claims and commitments, then the provider can be trusted. Although these studies were designed to understand how consumers evaluate peer-to-peer services, the findings about how uncertainty affects trust assessments and ultimately ratings, could likely be applied in many commercial contexts

where uncertainty is expected to higher, such as new product innovations, or credence goods like car repair.

The relationship between expectancy disconfirmation and trust may operate differently when expectations are negatively disconfirmed. Trust assessments include perceptions of reliability and integrity. Study 5 demonstrates that when expectations are negatively disconfirmed, the locus of causality for the service failure affects the consumer's perception of the provider's reliability. Study 5 further demonstrates that when expectations are negatively disconfirmed, a provider will be considered to lack integrity only if the provider is deemed to have control over the performance failure. If so, and if the provider lacks integrity, then this may ultimately lead consumers to post lower ratings for their provider. This finding was further tested in Study 6.

The second contribution is to demonstrate the different causes of the positivity bias in peer-to-peer ratings. I demonstrate that contextual factors that are unique to peer-to-peer services contribute to the ratings bias through network effects and social norms. These biases are different than the self-selection bias (driven by satisfaction) that leads to the j-shaped ratings distribution in many commercial products and services (Schoenmüller et al., 2018). The Pretest Study confirms extant research suggesting that the review rate is much higher in peer-to-peer services (Fradkin, 2017). Thus, consumers with moderate opinions are more likely to be included in the peer-to-peer compared to the commercial ratings distribution. Each of the studies further confirms that expectations do not

significantly differ between peer-to-peer and commercial services and that satisfaction is not the main driver of the consistently positive ratings in peer-to-peer services.

Table 2 provides a list of proposed contextual differences in peer-to-peer services that could potentially bias the ratings for peer providers. Each of these differences were mostly (but not always) supported in the package of studies. Among these, the Pretest Study shows that network-related contextual effects cause consumers to feel that their ratings are more important to peer providers than commercial providers, and to feel a higher need to justify their ratings to peer providers than commercial providers. The perceived importance of ratings is higher in peer-to-peer networks because providers rely on them as their only means of customer acquisition, and because the platforms reward and punish providers based on ratings. Thus consumers may not want to unnecessarily harm peer providers with low ratings. Relatedly, because most peer-to-peer ratings are two-sided, consumers feel a need to justify their ratings decisions to providers and may fear that providers can retaliate against them. Study 3 shows that these network effects can bias peer-to-peer ratings by weakening the effect of satisfaction (but not trust) on ratings and by motivating consumers to post higher overall ratings than may otherwise be deserved. This can lead to highly positive ratings even when a provider's performance merely met expectations and was only moderately satisfying.

Studies 5 and 6 showed that consumers of peer-to-peer services may give highly positive ratings even when their expectations are negatively disconfirmed (i.e., when the service experience is worse than they expected). Part of the reason for this is the fact that

provider causality and provider control may be more difficult to determine for peer-to-peer services, because providers are not professionals and the standards of evaluation are less clear. However, even if a peer provider is deemed to have caused a service failure, and is thus considered to be unreliable, a consumer may give the provider positive ratings because they feel the pressure of social norms. Peer-to-peer exchanges, which blend economic and social exchange (Sundararajan, 2019), lead to feelings of gratitude and empathy because peer providers have invited consumers to share their homes, personal items, and time and space with them (Albinsson & Yasanthi Perera, 2012; Hamari et al., 2016; Hellwig et al., 2015). These norms motivate consumers to forgive unreliable providers for product failure. That is, social norms weaken the effects of perceptions of unreliability on ratings. However, if a provider lacks integrity, social norms no longer apply. By prioritizing their own interest over the interests of the consumer, the provider has moved the relationship into a purely economic rather than a social exchange. The provider's lack of integrity gives consumers the justification to give lower ratings, regardless of social norms.

The third contribution is the development and demonstration of three different potential solutions that platforms could use to help attenuate the positive ratings bias for peer-to-peer services. First, to attenuate the bias from the network effects of perceived ratings importance and need to justify ratings, peer-to-peer platforms can help consumers to feel that their ratings are anonymous. Study 3 demonstrated that when ratings are aggregated, and believed to be anonymous, they are less important. Similarly, consumers won't feel the same need to justify their rating because it is not identifiable to the provider. This

leads to lower overall ratings which may be more reflective of a provider's true performance. Many platforms, such as Airbnb, already attempt to anonymize ratings. They do so by not releasing ratings to providers until an individual provider has been rated by multiple consumers. After a certain number of consumers have rated a provider, the aggregate rating is released to the provider, and also publicly on the platform website. However, when the next consumer makes a rating, the provider could see how the aggregate rating is changed, and may be able to infer whether the consumer rated them positively or poorly. Whether they do so or not, is not important. What is important is whether a consumer believes that the provider can determine their individual rating. In a separate survey of RVezy consumers, I asked participants to tell me whether they rated their provider completely honestly or whether their rating was inflated. If the rating was inflated, I asked them to explain why they made that choice. Participants indicated that they were worried that a provider may react to a low rating by giving the participant a low rating as retribution. But RVezy releases ratings to providers and consumers simultaneously, so in practice, ratings retribution is not possible. Again, it is the consumer's perception, rather than reality, that is important. RVezy and other platforms should do more to anonymize ratings and to communicate with their members in a way that clearly explains how their anonymity is protected.

The second way to attenuate ratings bias is to make standards of evaluation more clear in peer-to-peer services. Study 5 demonstrated that consumers of peer-to-peer services are relatively unclear about the standards against which they should evaluate their peer providers. This is a problem because a lack of clarity makes it difficult to determine

whether a provider was the cause of the consumer's missed expectations. If consumers are unclear about evaluation standards, it is difficult to make a definitive assessment that the provider failed. Consumers may instead blame themselves for missed expectations. Study 5 further showed that when a platform can more clearly define the performance standards for peer-to-peer services, consumers can more easily identify performance failure and hold providers accountable for that failure. Ratings will be lower, which may be more reflective of a provider's true performance. However, as further demonstrated, social norms may weaken the effect of unreliability, and so ratings may still be somewhat biased. To completely remove these biases, platforms need to make it more easy to assess both causality and controllability, because it will affect perceptions of provider integrity.

Study 6 attempted to increase provider causality and controllability through a novel ratings system that represents the third way that platforms could attenuate the positive ratings bias. This new system puts the onus on providers to rate themselves on several attributes. Consumers can use their ratings to help select a provider, and can also base their expectations on the ratings. Providers will be expected to rate themselves honestly, and if they fail to meet those standards, consumers should feel that the provider caused and controlled the negative outcome. Study 6 demonstrates that when ratings are provider-authored, and especially if they are all five-stars, negative disconfirmation leads to perceptions that the provider lacks integrity. Ratings will be lower, which may be more reflective of a provider's true performance. There are several other potential advantages to this new rating system. First, it should help to differentiate quality, because providers will be incented to provide accurate ratings on all attributes. The best providers are able

stand out, based on their higher attribute ratings. Second, assuming that the ratings are accurate, future guests should have better experiences and higher satisfaction, because experiences will more often meet expectations. Third, the proposed system may encourage providers to improve their quality over time. In the current system, when providers learn that consumers will give a high rating even for mediocre quality, they may decide to cut back their effort. However, in the new system, providers may want to improve quality so that they can honestly give themselves a higher rating for each attribute. In sum, the proposed ratings system should help to fix the positive bias, leading to more satisfactory experiences for consumers, and a desire to remain in the platform to purchase additional service experiences in future.

11.2 Limitations and Future Directions

I tested my propositions in online and lab studies, and in a field experiment with consumers of a peer-to-peer service for recreational vehicles. In each case, I endeavored to promote realism in the studies. To properly mimic the contextual factors in peer-to-peer services, it was important that participants believed that they were participating in a real service experience, and that their ratings mattered for their provider and for the business or platform. This is because the moderating factors of uncertainty, risk, network pressures, and social norms do not operate the same if participants know they are taking part in a study. Overall, I believe that I was successful in maintaining the deception, as demonstrated by the fact that 90% of participants in Study 2B felt somewhat to completely sure that the experience was real. However, the fact that some participants did

not believe the deception may have contributed to some of the non-significant results in the studies (for example the moderating effect of uncertainty on trust in Study 2A).

I learned that it is very difficult to elicit negative disconfirmation while trying to maintain the deception of a real service experience. For example in Study 4, the manipulation was not heavy-handed enough in creating a poor product experience with the free cookie. I didn't want to arouse suspicion with a product that was objectively very bad. Thus I was not able to test most of the hypotheses in that study. However, I was able to successfully elicit negative disconfirmation in Study 5, which then allowed me to test all of the hypotheses. Similarly, the manipulations in Studies 2A and 2B did not drive significant differences in uncertainty and perceived risk. Again, I didn't want to arouse suspicion by making the privacy warnings too explicit. The level of uncertainty and perceived risk in studies 2A and 2B respectively were also relatively low overall which likely impacted the manipulation. The service experiences in those studies (resume design and graphic design) may have worked better if they were reversed.

The studies supported the majority of the propositions depicted in my conceptual models, but there were some unexpected results. For example, the results of Study 2B did not support the prediction that perceived risk moderates the effect of trust on ratings. A subsequent analysis in Study 2B and Study 3 demonstrated that perhaps perceived risk may indeed increase the importance of trust on the relationship with ratings, but through its effect on satisfaction rather than directly. When risk is high, and the consequences of performance failure are severe, consumers may be more satisfied when their expectations

are met, with no surprises. This finding should be tested in future research, and has implications for many categories both within peer-to-peer and commercial services in which perceived risk may be high. This includes, for example, services such as home renovations, e-commerce, and extreme sports.

Finally, I learned that a consumer's evaluations of peer-to-peer services may operate differently depending on the type of service and the level of sharing involved. I tested many different categories of services including food services, design services, recreational vehicle rental, and accommodation services. Perceptions of provider causality and controllability, and consumers feelings of gratitude and empathy may operate differently in peer-to-peer services that are skills-based, compared to those in which a peer provider shares their personal possessions and space with the consumer. These differences may have contributed to some of the non-significant results (for example, the fact that social norms of gratitude and empathy were not higher for peer-to-peer services compared to commercial services in Study 4). Future research should test these differences between different types of peer-to-peer services, and their effects on trust and ratings.

11.3 Final Thoughts

The peer-to-peer sharing economy is a large and growing industry, and one that is dependent on trust. However, the positive bias in peer-to-peer ratings makes it difficult to distinguish between providers, and may cause consumers to lose trust in the review

system and the platform itself. Indeed in Study 6, I find that the average peer-to-peer

rating is one point higher (on a five-star scale) than the objective quality of the service

(based on the ratings in a control condition). Study 6 was an online study that was not

able to mimic all of the real contextual effects of peer-to-peer platforms (for example it

did not include a two-sided review system), so the ratings bias may actually be even

higher. Despite these concerns, little progress has been made in demonstrating the cause

of the bias and how it can be fixed by the platforms. A notable exception is the research

on Airbnb by Fradkin (2017) which shows that the threat of retaliatory behavior is an

unintended consequence when consumer and provider ratings are not released

simultaneously. This unintended threat does not fully explain the positive bias (Fradkin,

2017), and most platforms (including Airbnb) now post ratings simultaneously. Therefore

the effects of overt retaliatory behavior on ratings have largely been removed (Bridges &

Vásquez, 2016). Thus, I do not consider these retaliatory effects in my studies.

Some researchers have proposed alternative causes for the positive ratings bias (e.g.,

Filippas et al., 2018; Mulshine, 2015) but they not have explicated the mechanisms of

their proposed causes. I attempted to address this issue by exploring how peer-to-peer

consumers evaluate their service experiences differently (i.e., through trust), and how this

difference, in concert with network and social factors that are unique to the peer-to-peer

context, lead to high ratings for peer providers. In doing so, I strive to make both a

theoretical contribution and a practical contribution, that includes possible solutions that

platforms could use to help attenuate the ratings bias.

***

12 References

Aggarwal, P. (2004). The effects of brand relationship norms on consumer attitudes and behavior. *Journal of Consumer Research*, *31*(1), 87–101.

Agustin, C., & Singh, J. (2005). Curvilinear Effects of Consumer Loyalty Determinants in Relational Exchanges. *Journal of Marketing Research*, *42*(1), 96–108.

Albinsson, P. A., & Yasanthi Perera, B. (2012). Alternative marketplaces in the 21st century: Building community through sharing events. *Journal of Consumer Behaviour*, *11*(4), 303–315. https://doi.org/10.1002/cb.1389

Anderson, E. W. (1998). Customer Satisfaction and Word of Mouth. *Journal of Service Research*, *1*(1), 5–17.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Anderson, E. W., & Sullivan, M. W. (1993). The antecedents and consequences of customer satisfaction for firms. *Marketing Science*, *12*(2), 125–143.

Athey, S., Castillo, J. C., & Knoepfle, D. (2018). *Service Quality in the Gig Economy*.

Avery, C., Resnick, P., & Zeckhauser, R. (1999). The Market for Evaluations. *The American Economic Review*, *89*(3), 564–584.

Awad, N. F., & Ragowsky, A. (2008). Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems*, *24*(4), 101–121.

Ba, S., & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS Quarterly*, 243–268.

Belk, R. (2013). Sharing versus pseudo-sharing in Web 2.0. *The Anthropologist*, *18*(1), 7–23.

Benoit, S., Baker, T. L., Bolton, R. N., Gruber, T., & Kandampully, J. (2017). A triadic framework for collaborative consumption (CC): Motives, activities and resources & capabilities of actors. *Journal of Business Research*, *79*, 219–227. https://doi.org/10.1016/j.jbusres.2017.05.004

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, *24*(4), 586–607. https://doi.org/10.1016/j.jcps.2014.05.002

Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of Management Review*, *23*(3), 459–472.

Blois, K. J. (1999). Trust in Business to Business Relationships: An Evaluation of its Status. *Journal of Management Studies*, *36*(2), 197–215.

Bolton, G., Greiner, B., & Ockenfels, A. (2012). Engineering Trust: Reciprocity in the Production of Reputation Information. *Management Science*, *59*(2), 265–285. https://doi.org/10.1287/mnsc.1120.1609

Botsman, R. (2013). *The sharing economy lacks a shared definition*. Fast Company. https://www.fastcompany.com/3022028/the-sharing-economy-lacks-a-shared-definition

Boulding, W., Kalra, A., Staelin, R., & Zeithaml, V. A. (1993). A dynamic process model of service quality: From expectations to behavioral intentions. *Journal of Marketing Research*, *30*(1), 7–27.

Bridges, J., & Vásquez, C. (2016). If nearly all Airbnb reviews are positive, does that make them meaningless? *Current Issues in Tourism*, *0*(0), 1–19. https://doi.org/10.1080/13683500.2016.1267113

Brown, T. J., Barry, T. E., Dacin, P. A., & Gunst, R. F. (2005). Spreading the word: Investigating antecedents of consumers' positive word-of-mouth intentions and behaviors in a retailing context. *Journal of the Academy of Marketing Science*, *33*(2), 123. https://doi.org/10.1177/0092070304268417

Caldieraro, F., Zhang, J. Z., Cunha Jr, M., & Shulman, J. D. (2018). Strategic information transmission in peer-to-peer lending markets. *Journal of Marketing*, *82*(2), 42–63.

Casado-Aranda, L.-A., Dimoka, A., & Sánchez-Fernández, J. (2019). Consumer processing of online trust signals: A neuroimaging study. *Journal of Interactive Marketing*, *47*, 159–180.

Celata, F., Hendrickson, C. Y., & Sanna, V. S. (2017). The sharing economy as community marketplace? Trust, reciprocity and belonging in peer-to-peer accommodation platforms. *Cambridge Journal of Regions, Economy and Society*, *10*(2), 349–363.

Cheema, A. (2008). Surcharges and Seller Reputation. *Journal of Consumer Research*, *35*(1), 167–177. https://doi.org/10.1086/529532

Chen, Y., & Chou, T. (2012). Exploring the continuance intentions of consumers for B2C online shopping: Perspectives of fairness and trust. *Online Information Review*.

Cheng, M. (2016). Sharing economy: A review and agenda for future research. *International Journal of Hospitality Management*, *57*, 60–70. https://doi.org/10.1016/j.ijhm.2016.06.003

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Chiou, J.-S., & Droge, C. (2006). Service quality, trust, specific asset investment, and expertise: Direct and indirect effects in a satisfaction-loyalty framework. *Journal of the Academy of Marketing Science*, *34*(4), 613–627.

Chopra, K., & Wallace, W. A. (2003). Trust in electronic environments. *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings.*, 10 pp.

Clark, M. S., & Mills, J. (2011). A Theory of Communal (and Exchange) Relationships. In *Handbook of Theories of Social Psychology: Volume Two*. SAGE.

Costello, J. P., & Reczek, R. W. (2020). Providers Versus Platforms: Marketing Communications in the Sharing Economy. *Journal of Marketing*, *84*(6), 22–38.

Curren, M. T., & Folkes, V. S. (1987). Attributional influences on consumers' desires to communicate about products. *Psychology & Marketing*, *4*(1), 31–45.

Darke, P. R., Ashworth, L., & Main, K. J. (2010). Great expectations and broken promises: Misleading claims, product failure, expectancy disconfirmation and consumer distrust. *Journal of the Academy of Marketing Science*, *38*(3), 347–362.

De Langhe, B., Fernbach, P. M., & Lichtenstein, D. R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, *42*(6), 817–833.

Dellarocas, C. (2003). The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, *49*(10), 1407–1424. https://doi.org/10.1287/mnsc.49.10.1407.17308

Eckhardt, G. M., Houston, M. B., Jiang, B., Lamberton, C., Rindfleisch, A., & Zervas, G. (2019). Marketing in the sharing economy. *Journal of Marketing*, *83*(5), 5–27.

Eisenbeiss, M., Cornelißen, M., Backhaus, K., & Hoyer, W. D. (2014). Nonlinear and asymmetric returns on customer satisfaction: Do they vary across situations and consumers? *Journal of the Academy of Marketing Science*, *42*(3), 242–263.

Engler, T. H., Winter, P., & Schulz, M. (2015). Understanding online product ratings: A customer satisfaction model. *Journal of Retailing and Consumer Services*, *27*, 113–120.

Filippas, A., Horton, J. J., & Golden, J. (2018). Reputation inflation. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 483–484.

Folkes, V. S. (1988). Recent attribution research in consumer behavior: A review and new directions. *Journal of Consumer Research*, *14*(4), 548–565.

Fornell, C. (1992). A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, *56*(1), 6–21.

Fradkin, A. (2017). *Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb* (SSRN Scholarly Paper ID 2939084). Social Science Research Network. https://papers.ssrn.com/abstract=2939084

Fradkin, A., Grewal, E., Holtz, D., & Pearson, M. (2015). Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 641–641.

Ganesan, S. (1994). Determinants of long-term orientation in buyer-seller relationships. *Journal of Marketing*, *58*(2), 1–19.

Garbarino, E., & Johnson, M. S. (1999). The Different Roles of Satisfaction, Trust, and

Commitment in Customer Relationships. *Journal of Marketing*, *63*(2), 70–87.

https://doi.org/10.2307/1251946

Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth

communication. *Marketing Science*, *23*(4), 545–560.

Grant, S. J., & Tybout, A. M. (2008). The effect of temporal frame on information

considered in new product evaluation: The role of uncertainty. *Journal of

Consumer Research*, *34*(6), 897–913.

Gremler, D. D., Gwinner, K. P., & Brown, S. W. (2001). Generating positive word-of-

mouth communication through customer-employee relationships. *International

Journal of Service Industry Management*, *12*(1), 44–59.

Gullstrand Edbring, E., Lehner, M., & Mont, O. (2016). Exploring consumer attitudes to

alternative models of consumption: Motivations and barriers. *Journal of Cleaner

Production*, *123*, 5–15. https://doi.org/10.1016/j.jclepro.2015.10.107

Halstead, D., Hartman, D., & Schmidt, S. L. (1994). Multisource effects on the

satisfaction formation process. *Journal of the Academy of Marketing Science*,

22(2), 114–129.

Hamari, J., Sjöklint, M., & Ukkonen, A. (2016). The sharing economy: Why people

participate in collaborative consumption. *Journal of the Association for

Information Science and Technology*, *67*(9), 2047–2059.

https://doi.org/10.1002/asi.23552

Hamilton, R., Vohs, K. D., & McGill, A. L. (2014). We'll Be Honest, This Won't Be the

Best Article You'll Ever Read: The Use of Dispreferred Markers in Word-of-

Mouth Communication. *Journal of Consumer Research*, *41*(1), 197–212. https://doi.org/10.1086/675926

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second Edition). Guilford Publications.

Heider, F. (1958). *The psychology of interpersonal relations*. Psychology Press.

Hellwig, K., Belk, R. W., & Morhart, F. (2015). Shared Moments of Sociality: Embedded Sharing Within Peer to Peer Hospitality Platforms. *Advances in Consumer Research*, *43*, 186–190.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, *18*(1), 38–52.

Homburg, C., Koschate, N., & Hoyer, W. D. (2005). Do satisfied customers really pay more? A study of the relationship between customer satisfaction and willingness to pay. *Journal of Marketing*, *69*(2), 84–96.

Hong, J., & Chang, H. H. (2015). "I" follow my heart and "We" rely on reasons: The impact of self-construal on reliance on feelings versus reasons in decision making. *Journal of Consumer Research*, *41*(6), 1392–1411.

Horton, J., & Golden, J. (2015). Reputation inflation: Evidence from an online labor market. *Work. Pap., NYU*.

Hu, N., Pavlou, P. A., & Zhang, J. J. (2017). On Self-Selection Biases in Online Product Reviews. *MIS Quarterly*, *41*(2), 449–471.

Ikkala, T., & Lampinen, A. (2015). Monetizing Network Hospitality: Hospitality and Sociability in the Context of Airbnb. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1033–1044. https://doi.org/10.1145/2675133.2675274

Jenkins, A. (2018, April 5). *Giving Your Uber Driver 5 Stars Isn't Helping Anyone*. Fortune.Com. http://fortune.com/2018/04/05/uber-negative-ratings-stars/

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, *47*, 278.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence-versus integrity-based trust violations. *Journal of Applied Psychology*, *89*(1), 104.

Kim, Y., & Peterson, R. A. (2017). A Meta-analysis of Online Trust Relationships in E-commerce. *Journal of Interactive Marketing*, *38*, 44–54.

Lee, S., Winterich, K. P., & Ross Jr, W. T. (2014). I'm moral, but I won't help you: The distinct roles of empathy and justice in donations. *Journal of Consumer Research*, *41*(3), 678–696.

Lee, Y.-J., Hosanagar, K., & Tan, Y. (2015). Do I follow my friends or the crowd? Information cascades in online movie ratings. *Management Science*, *61*(9), 2241–2258.

Lewicki, R. J., & Bunker, B. B. (1995). Developing and Maintaining Trust in Work Relationships. In *Trust in Organizations: Frontiers of Theory and Research*. SAGE Publications.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022.

Lewis, J. D., & Weigert, A. (1985). Trust as a Social Reality. *Social Forces*, *63*(4), 967–985. https://doi.org/10.1093/sf/63.4.967

Lovett, M. J., Peres, R., & Shachar, R. (2013). On brands and word of mouth. *Journal of Marketing Research*, *50*(4), 427–444.

Luffarelli, J., Mukesh, M., & Mahmood, A. (2019). Let the logo do the talking: The influence of logo descriptiveness on brand equity. *Journal of Marketing Research*, *56*(5), 862–878.

Martin, S. S., Camarero, C., & José, R. S. (2011). Does involvement matter in online shopping satisfaction and trust? *Psychology & Marketing*, *28*(2), 145–167.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model Of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/AMR.1995.9508080335

Meuter, M. L., Ostrom, A. L., Roundtree, R. I., & Bitner, M. J. (2000). Self-service technologies: Understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, *64*(3), 50–64.

Mittal, V., Ross Jr, W. T., & Baldasare, P. M. (1998). The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, *62*(1), 33–47.

Moe, W. W., & Schweidel, D. A. (2011). Online Product Opinions: Incidence, Evaluation, and Evolution. *Marketing Science*, *31*(3), 372–386. https://doi.org/10.1287/mksc.1110.0662

Moe, W. W., & Trusov, M. (2011). The Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, *48*(3), 444–456. https://doi.org/10.1509/jmkr.48.3.444

Molm, L. D., Takahashi, N., & Peterson, G. (2000). Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition. *American Journal of Sociology*, *105*(5), 1396–1427. https://doi.org/10.1086/210434

Morales, A. C. (2005). Giving firms an "E" for effort: Consumer responses to high-effort firms. *Journal of Consumer Research*, *31*(4), 806–812.

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. *Journal of Marketing*, *58*(3), 20–38. https://doi.org/10.2307/1252308

Mulshine, M. (2015). *After a disappointing Airbnb stay, I realized there's a major flaw in the review system*. Business Insider. http://www.businessinsider.com/why-airbnb-reviews-are-a-problem-for-the-site-2015-6

Nosko, C., & Tadelis, S. (2015). *The Limits of Reputation in Platform Markets: An Empirical Analysis and Field Experiment* (Working Paper No. 20830). National Bureau of Economic Research. https://doi.org/10.3386/w20830

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469.

Oliver, R. L. (1993). Cognitive, Affective, and Attribute Bases of the Satisfaction Response. *Journal of Consumer Research*, *20*(3), 418–430.

Oliver, R. L. (2010). *Satisfaction: A behavioral perspective on the consumer: A behavioral perspective on the consumer*. Routledge.

Patterson, P. G., Johnson, L. W., & Spreng, R. A. (1996). Modeling the determinants of customer satisfaction for business-to-business professional services. *Journal of the Academy of Marketing Science*, *25*(1), 4–17.

Ranaweera, C., & Prabhu, J. (2003). On the relative importance of customer satisfaction and trust as determinants of customer retention and positive word of mouth. *Journal of Targeting, Measurement and Analysis for Marketing*, *12*(1), 82–90. https://doi.org/10.1057/palgrave.jt.5740100

Resnick, P., & Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system. In *The Economics of the Internet and E-commerce* (Vol. 11, pp. 127–157). Emerald Group Publishing Limited. https://doi.org/10.1016/S0278-0984(02)11030-3

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, 574–599.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/AMR.1998.926617

Schlosser, A. E. (2005). Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, *32*(2), 260–265.

Schoenmüller, V., Netzer, O., & Stahl, F. (2018). The extreme distribution of online reviews: Prevalence, drivers and implications. *Columbia Business School Research Paper*, *18–10*.

Schor, J. (2014). *Debating the sharing economy*. Great Transition Initiative.

Schurr, P. H., & Ozanne, J. L. (1985). Influences on exchange processes: Buyers'
preconceptions of a seller's trustworthiness and bargaining toughness. *Journal of Consumer Research*, *11*(4), 939–953.

Selnes, F. (1998). Antecedents and consequences of trust and satisfaction in buyer-seller
relationships. *European Journal of Marketing*, *32*(3/4), 305–322.
https://doi.org/10.1108/03090569810204580

Simonson, I., & Rosen, E. (2014). What marketers misunderstand about online reviews.
Harvard Business Review, *92*(1), 7.

Singh, J., & Sirdeshmukh, D. (2000). Agency and trust mechanisms in consumer
satisfaction and loyalty judgments. *Journal of the Academy of Marketing Science*,
28(1), 150–167.

Sirdeshmukh, D., Singh, J., & Sabol, B. (2002). Consumer Trust, Value, and Loyalty in
Relational Exchanges. *Journal of Marketing*, *66*(1), 15–37.
https://doi.org/10.1509/jmkg.66.1.15.18449

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the Determinants of Risk
Behavior. *Academy of Management Review*, *17*(1), 9–38.
https://doi.org/10.5465/AMR.1992.4279564

Slee, T. (2013). *Some obvious things about internet reputation systems*. http://tomslee.
net

Smith, J. B. (1998). Buyer–Seller relationships: Similarity, Relationship Management,
and Quality. *Psychology and Marketing*, *15*(1), 3–21.

Smith, J. B., & Barclay, D. W. (1997). The Effects of Organizational Differences and

    Trust on the Effectiveness of Selling Partner Relationships. *Journal of Marketing*,

    61(1), 3–21. https://doi.org/10.2307/1252186

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking

    intentions and perception of trust. *Tourism Management*, *32*(6), 1310–1323.

Spreng, R. A., MacKenzie, S. B., & Olshavsky, R. W. (1996). A Reexamination of the

    Determinants of Consumer Satisfaction. *Journal of Marketing*, *60*(3), 15–32.

    https://doi.org/10.1177/002224299606000302

Sridhar, S., & Srinivasan, R. (2012). Social Influence Effects in Online Product Ratings.

    Journal of Marketing, *76*(5), 70–88. https://doi.org/10.1509/jm.10.0377

Stofberg, N., Bridoux, F., Ciulli, F., Pisani, N., Kolk, A., & Vock, M. (2019). A

    relational-models view to explain peer-to-peer sharing. *Journal of Management*

    Studies.

Sun, J., Keh, H. T., & Lee, A. Y. (2012). The effect of attribute alignability on service

    evaluation: The moderating role of uncertainty. *Journal of Consumer Research*,

    39(4), 831–847.

Sundararajan, A. (2019). Commentary: The twilight of brand and consumerism? Digital

    trust, cultural meaning, and the quest for connection in the sharing economy.

    Journal of Marketing, *83*(5), 32–35.

Ter Huurne, M., Ronteltap, A., Corten, R., & Buskens, V. (2017). Antecedents of trust in

    the sharing economy: A systematic review. *Journal of Consumer Behaviour*,

    16(6), 485–498.

Tsiros, M., Mittal, V., & Ross Jr, W. T. (2004). The role of attributions in customer satisfaction: A reexamination. *Journal of Consumer Research*, *31*(2), 476–483.

Tussyadiah, I. P. (2016). Factors of satisfaction and intention to use peer-to-peer accommodation. *International Journal of Hospitality Management*, *55*, 70–80. https://doi.org/10.1016/j.ijhm.2016.03.005

Utz, S., Kerkhof, P., & Van Den Bos, J. (2012). Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research and Applications*, *11*(1), 49–58.

Vlachos, P. A., Vrechopoulos, A. P., & Pramatari, K. (2011). Too much of a good thing: Curvilinear effects in the evaluation of services and the mediating role of trust. *Journal of Services Marketing*, *25*(6), 440–450.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548.

Weiner, B. (1995). *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. Guilford Press.

Woodruff, R. B., Cadotte, E. R., & Jenkins, R. L. (1983). Modeling Consumer Satisfaction Processes Using Experience-Based Norms. *Journal of Marketing Research*, *20*(3), 296–304. JSTOR. https://doi.org/10.2307/3151833

Zervas, G., Proserpio, D., & Byers, J. (2015). *A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average* (SSRN Scholarly Paper ID 2554500). Social Science Research Network. https://papers.ssrn.com/abstract=2554500

Zhang, Q., Yang, X. J., & Robert, L. P. (2020). *Expectations and trust in automated vehicles*. 1–9.

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths

    and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–

    206.

Zhu, F., & Zhang, X. (Michael). (2010). Impact of Online Consumer Reviews on Sales:

    The Moderating Role of Product and Consumer Characteristics. *Journal of*

    Marketing, *74*(2), 133–148. https://doi.org/10.1509/jm.74.2.133

Appendices

Appendix A: List of Propositions

P1   Consumers rate their providers based on an evaluation of perceived quality compared to initial expectations. For commercial services, the satisfaction that results from this evaluation is the primary driver of ratings. In contrast, for peer-to-peer services, expectancy disconfirmation evaluations affect ratings outside of satisfaction.

P2   If a consumer is relatively uncertain that a provider is able and willing to deliver positive outcomes as expected, the provider's fulfillment of those expectations has a stronger impact on the consumer's perceptions of trust than if they were relatively certain about outcomes.

P3   If an exchange carries a relatively high perceived risk, the provider's trustworthiness is an important performance attribute which should be reflected in their rating.

P4   Peer-to-peer services ratings will reflect the consumer's trust in the peer provider, in addition to the consumer's satisfaction with the experience.

P5   When expectations are confirmed (i.e. merely met), trust will be higher than satisfaction.

P6a   Because satisfaction is relatively subjective, the effect of satisfaction on ratings will be reduced when consumers feel that ratings are very important to peer providers and when they have a strong need to justify their ratings.

P6b   When ratings are considered to be very important to their providers, consumers will post higher ratings for their provider than they post when ratings are considered less important.

P7   If a consumer is not certain that a provider caused a performance failure, then the effect of negative disconfirmation on perceptions of provider reliability is weakened.

P8   If a consumer is not certain that a provider had control over a performance failure, then the effect of negative disconfirmation on perceptions of provider integrity is weakened.

P9a   When gratitude and empathy are high, consumers will post higher ratings for their provider than when gratitude and empathy are lower.

P9b   Gratitude and empathy will cause consumers to forgive a provider for unreliable service, but not if the provider is deemed to lack integrity. Thus, when negative disconfirmation occurs, gratitude and empathy weaken the effect of reliability on ratings but not the effect of integrity on ratings.

Appendix B: Examples of Airbnb Accommodations

Rent an entire Castle (sleeps 16) & Grounds
★ 4.86 (120 reviews) · Superhost · West Ashton, England, United Kingdom

Opulent Toronto Mansion with Pool
★ 5.0 (17 reviews) · Toronto, Ontario, Canada

The Guide Tent
★ 4.95 (73 reviews) · Superhost · Whitney, Ontario, Canada

Muskoka Treehouse
★ 4.95 (37 reviews) · Superhost · Minett, Ontario, Canada

Small studio
★ 4.83 (126 reviews) · Superhost · Brampton, Ontario, Canada
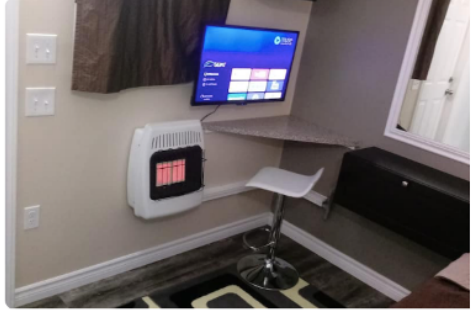
A School Bus on the Cabot Trail
★ 4.82 (28 reviews) · Superhost · Englishtown, Nova Scotia, Canada

grounding self contained basement in the east
★ 4.94 (18 reviews) · Superhost · Scarborough, Ontario, Canada

Female Dorm Room-Pearl
★ 3.83 (6 reviews) · Toronto, Ontario, Canada

Appendix C: Summary Statistics and Correlation Matrix for Pretest Study

Correlation Matrix

	Mean	SD	α	1	2	3	4	5	6	7	8	9	10	11	12
1 Expectations	4.52	1.25		1											
2 Perceived Risk	2.68	1.58		-.06	1										
3 Uncertainty	2.94	1.15	.74	-.47**	.33**	1									
4 Satisfaction	5.76	1.45		.15*	-.21**	-.12	1								
5 Expectancy Disc.	4.55	1.35	.74	.170*	-.07	-.02	.60**	1							
6 Provider Causality	4.67	1.20	.74	-.15	-.24	.10	-.50**	-.13	1						
7 Provider Control	4.91	1.43	.94	.21	-.44	-.36	-.31	-.39	.45	1					
8 Rating	4.50	.90		.08	-.14	.06	.78**	.65**	-.43	-.59	1				
9 Ratings Importance	5.63	1.05	.80	.18*	.02	-.08	.26**	.26**	-.30	-.24	.23*	1			
10 Need to Justify	4.15	1.23	.75	.05	.19**	.05	-.10	-.02	.11	.04	-.16	.23**	1		
11 Clarity of Standards	4.85	1.50		.35**	-.08	-.33**	.03	.11	.24	.36	-.02	.12	.04	1	
12 Social Norms	2.89	.93	.90	.29**	-.09	-.18*	.49**	.54**	-.13	-.52*	.47**	.28**	.01	.14	1

** Correlation is significant at $p < .01$; * Correlation is significant at $p < .05$

Appendix D: Stimuli for Study 1

Commercial Business Condition

Peer-to-Peer Service Condition

Appendix E: Summary Statistics and Correlation Matrix for Study 1

		Mean	SD	Correlation Matrix			
				1	2	3	4
1	Expectations	6.12	1.76	1			
2	Rating	4.26	.70	.35**	1		
3	Satisfaction	8.48	1.35	.42**	.78**	1	
4	Expectancy Disc.	1.13	.71	-.13	.44**	.51**	1

** Correlation is significant at $p < .01$

Appendix F: Email from Graphic Designer to Participants in Study 2A

New + Commercial Condition

Appendix F Continued: Email from Graphic Designer to Participants in Study 2A

Established + Commercial Condition

Congratulations! Netwrk has assigned a graphic designer that is ready to take on your project!

Netwrk is a graphic design service that guarantees that its designers will deliver exceptional designs to clients within 24 hours.

Hi, I'm Michael!

I'm a graphic designer specializing in Photoshop, Illustrator, and CorelDraw. Since I was young, I've had a love of all forms of art and design. I do my work with great passion and dedication, and I strive to deliver imaginative service. I've completed over 100 projects with Netwrk, and I'm excited to make all of your design wishes come true!

Hi,

This is Michael. I'm really happy to be working with you to create your new resume. To begin, I just have a few questions, to help

New + Peer-to-Peer Condition

Congratulations! Netwrk has matched you with a freelance graphic designer that is ready to take on your project!

Netwrk is a new peer-to-peer service that guarantees that freelance designers who have joined the platform will deliver exceptional designs to clients within 24 hours.
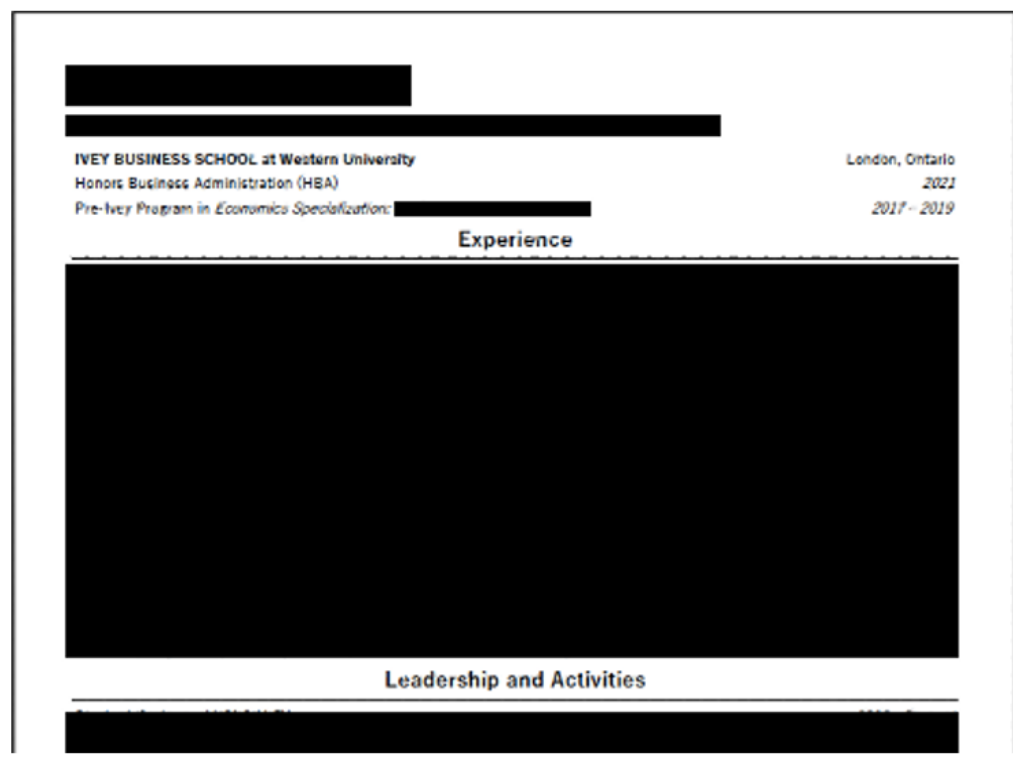
Hi, I'm Michael!

I'm a freelance graphic designer specializing in Photoshop, Illustrator, and CorelDraw. Since I was young, I've had a love of all forms of art and design. I do my work with great passion and dedication, and I strive to deliver imaginative service. I recently joined the Netwrk platform, and I'm excited to make all of your design wishes come true!

Hi,

This is Michael. I'm really happy to be working with you on one of my first projects to create your new resume. To begin, I just

Established + Peer-to-Peer Condition

Congratulations! Netwrk has matched you with a freelance graphic designer that is ready to take on your project!

Netwrk is a peer-to-peer service that guarantees that freelance designers on the platform will deliver exceptional designs to clients within 24 hours.

Hi, I'm Michael!

I'm a freelance graphic designer specializing in Photoshop, Illustrator, and CorelDraw. Since I was young, I've had a love of all forms of art and design. I do my work with great passion and dedication, and I strive to deliver imaginative service. I've completed over 100 projects with the Netwrk platform, and I'm excited to make all of your design wishes come true!
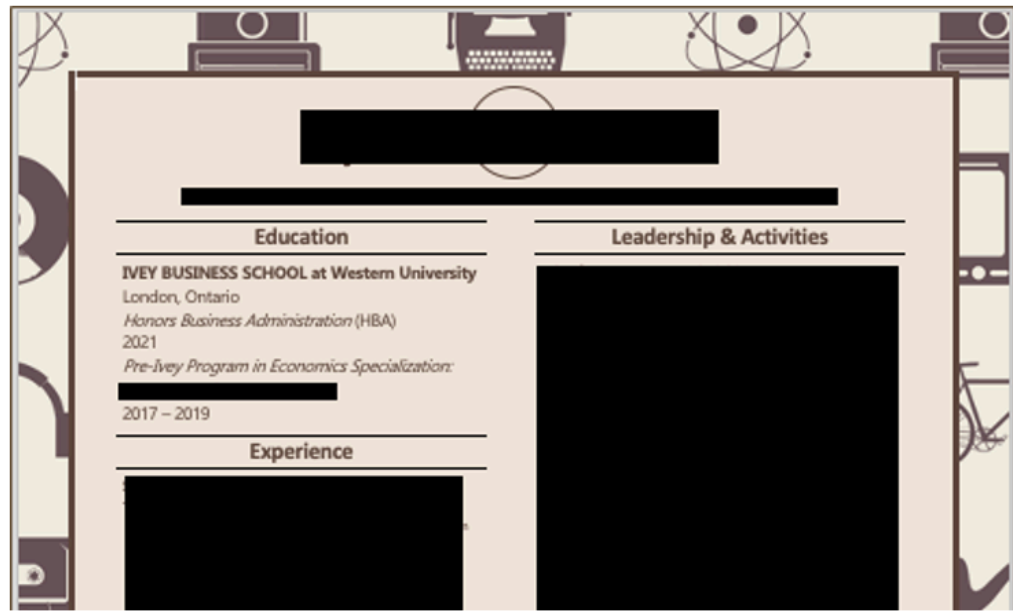
Hi,

This is Michael. I'm really happy to be working with you to create your new resume. To begin, I just have a few

Appendix G: Example Resume Designs for Study 2A

Original Resume from Participant 1

IVEY BUSINESS SCHOOL at Western University — London, Ontario
Honors Business Administration (HBA) — 2021
Pre-Ivey Program in *Economics Specialization:* — 2017 - 2019

Experience

Leadership and Activities

New Resume Design sent to Participant 1

Education

IVEY BUSINESS SCHOOL at Western University
London, Ontario
Honors Business Administration (HBA)
2021
Pre-Ivey Program in Economics Specialization:

2017 – 2019

Leadership & Activities

Experience

Appendix H: Summary Statistics and Correlation Matrix for Study 2A

		Mean	SD	α	Correlation Matrix						
					1	2	3	4	5	6	7
1	Expectations	5.39	1.06		1						
2	Perceived Risk	2.79	1.54		.11	1					
3	Uncertainty	2.86	1.03	.70	-.44**	.09	1				
4	Rating	4.10	.98		.03	-.08	-.23**	1			
5	Expectancy Disc.	4.35	1.67		-.00	-.12	-.19**	.71**	1		
6	Satisfaction	4.96	1.45		.05	-.16*	-.32**	.74**	.82**	1	
7	Trust	5.46	1.16	.90	.17*	-.17*	-.36**	.67**	.63**	.75**	1

** Correlation is significant at $p < .01$

* Correlation is significant at $p < .05$

Appendix I: Example Logo Designs for Study 2B

Appendix J: Summary Statistics and Correlation Matrix for Study 2B

		Mean	SD	α	Correlation Matrix						
					1	2	3	4	5	6	7
1	Expectations	5.03	1.18		1						
2	Perceived Risk	2.93	1.39		-.22**	1					
3	Uncertainty	3.03	1.04	.75	-.54**	.18*	1				
4	Rating	4.11	.96		.04	-.07	-.19*	1			
5	Expectancy Disc.	4.53	1.53		-.14	-.03	-.03	.73**	1		
6	Satisfaction	5.18	1.48		.05	-.06	-.26**	.83**	.74**	1	
7	Trust	5.74	1.17	.92	.06	-.04	-.33**	.72**	.68**	.78**	1

** Correlation is significant at $p < .01$
* Correlation is significant at $p < .05$

Appendix K: Reviews and Ratings on the RVezy.com Website (Study 3)

Appendix L: Summary Statistics and Correlation Matrix for Study 3

Correlation Matrix

		Mean	SD	α	1	2	3	4	5	6	7	8	9	10
1	Expectations	6.02	.95		1									
2	Perceived Risk	3.28	1.52		-.15**	1								
3	Uncertainty	2.31	1.02	.87	-.47**	.46**	1							
4	Rating	4.66	.68		.21**	-.01	-.19**	1						
5	Need to Justify	3.71	1.38	.73	-.06	.17**	.15**	-.20**	1					
6	Ratings Importance	6.07	.98	.80	.21**	-.08	-.30**	.46**	-.03	1				
7	Expectancy Disc.	5.83	1.30		.15**	.01	-.19**	.67**	-.28**	.37**	1			
8	Satisfaction	6.38	1.02		.23**	-.02	-.22**	.78**	-.21**	.40**	.67**	1		
9	Trust	6.71	.72	.95	.17**	-.01	-.20**	.63**	-.14**	.38**	.54**	.69**	1	
10	Trustworthiness	6.73	.71		.12*	-.06	-.17**	.58**	-.15**	.35**	.49**	.61**	.69**	1

** Correlation is significant at $p < .01$; * Correlation is significant at $p < .05$

Appendix M: Website Stimuli in Study 4

Appendix N: Summary Statistics and Correlation Matrix for Study 4

Correlation Matrix

	Mean	SD	α	1	2	3	4	5	6	7	8
1 Rating	4.74	.49		1							
2 Expectancy Disc.	3.99	.84		.30**	1						
3 Provider Control	3.00	1.43	.92	-.08	.b	1					
4 Satisfaction	5.99	1.24	.87	.26**	.35**	-.87	1				
5 Trust	4.52	.62	.87	.33**	.32**	.10	.47**	1			
6 Reliability	4.54	.66	.76	.35**	.31**	-.04	.45**	.94**	1		
7 Integrity	4.50	.67	.78	.28**	.29**	.20	.44**	.94**	.78**	1	
8 Social Norms	3.81	.87	.87	.28**	.44**	-.42	.47**	.36**	.35**	.32**	1

** Correlation is significant at $p < .01$

b. Cannot be computed because one of the variables is constant

Appendix O: Accommodation Options in Study 5

Commercial Business Conditions

Peer-to-Peer Service Conditions

Appendix P: Photos of Rental Accommodation Experience in Study 5

Appendix Q: Summary Statistics and Correlation Matrix for Study 5

Correlation Matrix

		Mean	SD	α	1	2	3	4	5	6	7	8	9	10
1	Rating	3.02	.89		1									
2	Expectancy Disc.	2.78	1.15		.74**	1								
3	Provider Causality	5.08	1.11	.73	-.33**	-.26**	1							
4	Provider Control	5.55	1.17	.89	-.13	-.24**	.33**	1						
5	Satisfaction	3.54	1.54		.81**	.76**	-.37**	-.20**	1					
6	Trust	3.78	1.24	.94	.63**	.66**	-.38**	-.48**	.68**	1				
7	Reliability	3.79	1.30	.89	.63**	.65**	-.39**	-.44**	.67**	.97**	1			
8	Integrity	3.77	1.27	.89	.60**	.61**	-.33**	-.46**	.64**	.97**	.87**	1		
9	Social Norms	2.00	.88	.93	.58**	.52**	-.30**	-.15*	.61**	.65**	.62**	.64**	1	
10	Clarity of Standards	4.88	1.45		.04	.04	.21**	.16*	.01	.03	.04	.02	.07	1

** Correlation is significant at $p < .01$; * Correlation is significant at $p < .05$

Appendix R: Examples of Designer Choices in Study 6

Condition 1
Consumer-authored
Five-star ratings
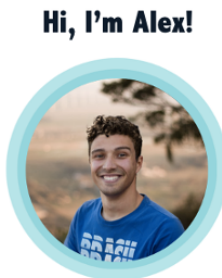
Condition 2
Provider-authored
Five-star ratings

Condition 3
Consumer-authored
Mixed ratings

Condition 4
Provider-authored
Mixed ratings

Appendix S: Examples of Logo Designs in Study 6

Appendix T: Summary Statistics and Correlation Matrix for Study 6

Correlation Matrix

	Mean	SD	α	1	2	3	4	5	6	7	8	9	10
1 Expectations	5.33	.97		1									
2 Perceived Risk	2.83	1.34		-.22**	1								
3 Uncertainty	2.73	.99	.71	-.37**	.40**	1							
4 Rating	4.13	.94		.05	-.14	-.25**	1						
5 Expectancy Disc.	4.38	1.55		.05	-.12	-.17*	.72**	1					
6 Provider Control	5.13	1.30	.90	-.17	.06	-.11	-.39*	-.47**	1				
7 Satisfaction	5.07	1.48		.06	-.24**	-.28**	.75**	.74**	.24	1			
8 Trust	5.74	1.07	.91	.02	-.18*	-.31**	.78**	.71**	-.46*	.76**	1		
9 Reliability	5.74	1.17	.86	.01	-.20**	-.31**	.78**	.70**	-.44*	.76**	.96**	1	
10 Integrity	5.74	1.07	.78	.03	-.14	-.29**	.72**	.66**	-.39*	.70**	.95**	.84**	1

** Correlation is significant at $p < .01$; * Correlation is significant at $p < .05$

Appendix U: Summary of Results of Hypothesis Tests

Study	Hypothesis	Supported?
1	1.1 For commercial services, the relationship between expectancy disconfirmation and ratings is fully mediated by satisfaction	Yes
1	1.2 For peer-to-peer services, expectancy disconfirmation has both a direct effect and an indirect effect through satisfaction on ratings	Yes
2A	2.1 Expectancy disconfirmation is positively related to trust in the service provider	Yes
2A, 2B, and 3	2.2 Uncertainty moderates the relationship between expectancy disconfirmation and trust; this relationship will be stronger when uncertainty is higher.	Unclear
2A, 2B	2.3 Trust is positively related to ratings	Yes
2B	2.4 Perceived risk moderates the relationship between trust and ratings; this relationship will be stronger when perceived risk is higher.	No
2B	2.5 When uncertainty and perceived risk are high, expectancy disconfirmation is positively related to ratings, and is mediated by trust in the service provider.	Unclear
3	3.1 Satisfaction is positively related to ratings.	Yes
3	3.2 A need to justify ratings decisions moderates the relationship between satisfaction and ratings; this relationship will be weaker when need to justify is higher.	Unclear
3	3.3 Perceived ratings importance moderates the relationship between satisfaction and consumer ratings; this relationship will be weaker when ratings importance is higher.	Yes
3	3.4 Perceived ratings importance is positively related to ratings.	Yes
2B, 3	3.5 Trust is positively related to satisfaction.	Yes
2B, 3	3.6 Perceived risk moderates the relationship between trust and satisfaction; this relationship will be stronger when perceived risk is higher.	Unclear

Appendix U Continued: Summary of Results of Hypothesis Tests

Study	Hypothesis	Supported?
5	5.1 Clarity of standards is positively related to perceptions of provider causality	Yes
5	5.2 Negative disconfirmation is negatively related to perceptions of provider reliability	Yes
5	5.3 Negative disconfirmation is negatively related to perceptions of provider integrity	Yes
5	5.4 Provider causality moderates the relationship between negative disconfirmation and reliability; this relationship will be stronger when provider causality is higher	No
5	5.5 Provider control moderates the relationship between negative disconfirmation and integrity; this relationship will be stronger when provider control is higher	Yes
4, 5	5.6 Social norms of gratitude and empathy are positively related to ratings.	Yes
5	5.7 When expectations are negatively disconfirmed, social norms moderate the relationship between reliability and ratings; this relationship will be weaker when social norms are higher	Yes
6	6.1 Compared to consumer-authored ratings, provider-authored ratings lead to higher perceived provider control	Unclear
6	6.2 The rating score moderates the relationship between provider-authored ratings and provider control; this relationship is stronger when the rating score is five-stars	Unclear
6	6.3 When the rating score is five-stars, provider-authored ratings lead to lower perceptions of provider integrity compared to when ratings are consumer-authored	Yes
6	6.4 When the rating score is five-stars, provider-authored ratings lead to lower post-experience ratings compared to when pre-experience ratings are consumer-authored	Yes

Appendix V: Ethics Approval Forms

Ethics Approval for Pretest Study

Western Research

Date: 19 December 2019

To: Dr. June Cotte

Project ID: 115151

Study Title: Consumer evaluations of peer and commercial services

Short Title: Consumer evaluations of peer and commercial services

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: January 10 2020

Date Approval Issued: 19/Dec/2019

REB Approval Expiry Date: 19/Dec/2020

Dear Dr. June Cotte

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
LOI with CONSENT for P2P study- version DEC-07-2019	Implied Consent/Assent	07/Dec/2019	
Recruitment Notice - Version Dec 7 2019	Recruitment Materials	07/Dec/2019	
Study Instrument - Peer and Commercial Services - version DEC-07-2019	Online Survey	07/Dec/2019	

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Kelly Patterson, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Appendix V Continued: Ethics Approval Forms

Ethics Approval for Studies 1, 3, and 4

Western Research

Date: 14 June 2018

To: Dr. June Cotte

Project ID: 111514

Study Title: Causes and Effects of Positivity Bias in Peer-to-Peer user Reviews

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 06/Jul/2018

Date Approval Issued: 14/Jun/2018 15:40

REB Approval Expiry Date: 14/Jun/2019

Dear Dr. June Cotte

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
CLEAN - Debriefing Document - Laboratory Study 1 - Version June 6 2018	Debriefing document	06/Jun/2018	2
CLEAN - LOI and Consent - Laboratory Study 1 - Version June 13 2018	Written Consent/Assent	13/Jun/2018	2
CLEAN - LOI and Consent - Online Study 2 - Version June 6 2018	Implied Consent/Assent	06/Jun/2018	2
CLEAN - Study 2 - Online Field Study Instrument - Version June 6 2018	Online Survey	06/Jun/2018	2
Debriefing Document - Online Study 2 - Version June 6 2018	Debriefing document	06/Jun/2018	
Recruitment Script - Laboratory Study 1 - April 25 2018	Oral Script	25/Apr/2018	1
STUDY 1 - Laboratory Study Instrument - April 25 2018	Paper Survey	25/Apr/2018	1
Study 2 – Recruitment Email - April 25 2018	Recruitment Materials	25/Apr/2018	1

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Katelyn Harris, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Appendix V Continued: Ethics Approval Forms

Ethics Approval for Studies 2A and 2B

Western Research

Date: 10 December 2019

To: Dr. June Cotte

Project ID: 114599

Study Title: Consumer evaluations of graphic design service

Application Type: NMREB Initial Application

Review Type: Delegated

Meeting Date: 04/Oct/2019 12:30

Date Approval Issued: 10/Dec/2019 12:52

REB Approval Expiry Date: 10/Dec/2020

_____

Dear Dr. June Cotte

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
114599 - Email with Link to survey 2 of 2 for Study 1A - version 15-08-2019	Recruitment Materials	15/Aug/2019	1
114599 - Email with Link to survey 2 of 2 for Study 1B - version 15-08-2019	Recruitment Materials	15/Aug/2019	1
114599 - SONA recruitment script for both Study 1A and Study 1B- version 15-08-2019	Recruitment Materials	15/Aug/2019	1
114599 - Study Instrument 2 of 2 for Study 1A - version 15-08-2019	Online Survey	15/Aug/2019	1
114599 - Study Instrument 2 of 2 for Study 1B - version 15-08-2019	Online Survey	15/Aug/2019	1
CLEAN - 114599 - Debriefing Document for both Study 1A and Study 1B - version 26-11-2019	Debriefing document	26/Nov/2019	3
CLEAN - 114599 - Email with Link to survey 1 of 2 for Study 1A - version 26-11-2019	Recruitment Materials	26/Nov/2019	3
CLEAN - 114599 - Email with Link to survey 1 of 2 for Study 1B - version 26-11-2019	Recruitment Materials	26/Nov/2019	3
CLEAN - 114599 - Study Instrument 1 of 2 for Study 1A - version 30-10-2019	Online Survey	30/Oct/2019	2
CLEAN - 114599 - Study Instrument 1 of 2 for Study 1B- version 30-10-2019	Online Survey	30/Oct/2019	2
CLEAN - LOI with CONSENT for both Study 1A and Study 1B - version 26-11-2019	Implied Consent/Assent	26/Nov/2019	2
Deceptive upfront LOI with CONSENT for both Study 1A and Study 1B - version 26-11-2019	Implied Consent/Assent	26/Nov/2019	1

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario

Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Katelyn Harris, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Appendix V Continued: Ethics Approval Forms

Ethics Approval for Study 5

Western Research

Date: 22 October 2020

To: Dr. June Cotte

Project ID: 115175

Study Title: Consumer evaluations of hospitality services

Short Title: Consumer evaluations of hospitality services

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: November 6 2020

Date Approval Issued: 22/Oct/2020

REB Approval Expiry Date: 22/Oct/2021

_____

Dear Dr. June Cotte

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
PRETEST Study Instrument for Hospitality Services - version OCT 2020	Online Survey	02/Oct/2020	
Online Study Instrument for Hospitality Services - version OCT 2 2020	Online Survey	02/Oct/2020	
Recruitment Notice Hospitality Study - Version OCT 2 2020	Recruitment Materials	02/Oct/2020	
LOI with CONSENT for hospitality study- version OCT 2-2020	Implied Consent/Assent	02/Oct/2020	
LOI with CONSENT for PRETEST hospitality study- version OCT 2-2020	Implied Consent/Assent	02/Oct/2020	

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Kelly Patterson, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Appendix V Continued: Ethics Approval Forms

Ethics Approval for Study 6

Date: 26 February 2020

To: Dr. June Cotte

Project ID: 115163

Study Title: Consumer evaluations of resume design service

Application Type: NMREB Initial Application

Review Type: Delegated

Full Board Reporting Date: 06/Mar/2020

Date Approval Issued: 26/Feb/2020 09:50

REB Approval Expiry Date: 26/Feb/2021

_____

Dear Dr. June Cotte

The Western University Non-Medical Research Ethics Board (NMREB) has reviewed and approved the WREM application form for the above mentioned study, as of the date noted above. NMREB approval for this study remains valid until the expiry date noted above, conditional to timely submission and acceptance of NMREB Continuing Ethics Review.

This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
CLEAN VERSION - SONA recruitment script for Resume Design Study - Version FEB-09-2020	Recruitment Materials	09/Feb/2020	
Debriefing Document for Resume Design Study - version DEC-18-2019	Debriefing Letter	18/Dec/2019	
Email with Link to survey 1 of 2 for resume design study - version DEC-18-2019	Recruitment Materials	18/Dec/2019	
Email with Link to survey 2 of 2 for Resume Design Study - version DEC-18-2019	Recruitment Materials	18/Dec/2019	
LOI with CONSENT for Resume design study- version DEC-18-2019	Implied Consent/Assent	18/Dec/2019	
Study Instrument 1 of 2 for Re´sume´ study - version DEC-18-2019	Online Survey	18/Dec/2019	
Study Instrument 2 of 2 for Re´sume´ Study - version DEC-18-2019	Online Survey	18/Dec/2019	

No deviations from, or changes to the protocol should be initiated without prior written approval from the NMREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Katelyn Harris, Research Ethics Officer on behalf of Dr. Randal Graham, NMREB Chair

Curriculum Vitae

Name:	Michael Moorhouse
Post-secondary Education and Degrees:	Wilfrid Laurier University Waterloo, Ontario, Canada 1997-2001 B.B.A. Ted Rogers School of Management, Ryerson University Toronto, Ontario, Canada 2011-2012 M.B.A. Ivey Business School, Western University London, Ontario, Canada 2016-2021 Ph.D.
Related Work Experience:	Instructor, Marketing Management Ivey Business School. 2020 Instructor, Principles of Marketing Western University, 2019 Sales and Marketing Director Nickelodeon Viacom Consumer Products, 2012-2016 Senior Category Marketing Manager Walmart Canada, 2009-2011 Brand Manager Procter & Gamble, 2004-2008
Honours and Awards:	AMA-Sheth Foundation Doctoral Consortium Fellow 2020-2021 Province of Ontario Graduate Scholarship 2018-2019, 2019-2020 Dean's Scholarship, Ivey Business School 2016-2020 Graduate Scholarship, Ted Rogers School of Management 2011-2012