UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

Creation And Validation of the Consumer Distrust in Specific Company Scale (CDISCS): A Novel Scale to Assess an Individual's Distrust for a Company

Permalink

https://escholarship.org/uc/item/1z1129zx

Author

Warshaw, Jeffrey Mark

Publication Date

2016

Supplemental Material

https://escholarship.org/uc/item/1z1129zx#supplemental

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, availalbe at https://creativecommons.org/licenses/by-nc-sa/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

CREATION AND VALIDATION OF THE CONSUMER DISTRUST IN SPECIFIC COMPANY SCALE (CDISCS): A NOVEL SCALE TO ASSESS AN INDIVIDUAL'S DISTRUST FOR A COMPANY

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Jeffrey Warshaw

December 2016

The Dissertation of Jeffrey Warshaw is approved:

_____
Professor Steve Whittaker, chair

_____
Professor Doug Bonett

_____
Professor Sri Kurniawan

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Jeffrey Warshaw

2016

Table of Contents

List of Figures

List of Tables

Abstract

Creation And Validation of the Consumer Distrust in Specific Company Scale (CDISCS): A Novel Scale to Assess an Individual's Distrust for a Company

Jeffrey Warshaw

Recent research has shown that distrust for a company is not merely the opposite of trust. Rather, it is a distinct construct that exists parallel to trust. Nevertheless, researchers studying distrust for a company have so far typically relied on reverse-worded trust scales which only apply to a single company. I created the Consumer Distrust in Specific Company Scale (CDISCS), a four-subscale instrument which can be deployed flexibly across most companies. Through three studies deploying the scale to participants on Mechanical Turk, I provide evidence for the current reliability and validity of the scale. I demonstrate the scale's high retest reliability, differential reliability across demographic comparisons, and ability to measure changes in consumer distrust in response to an authentic corporate scandal. I note that the internal consistency, inter-item correlations and inter-scale correlations of the Incompetence, Malevolence, and Intentions subscales are higher than desired, though within the same range as other trust and distrust scales considered validated in prior research. The CDISCS should therefore be considered equally ready to deploy as other published trust and distrust scales that have these features.

Acknowledgements

I wish to acknowledge and thank the following people for making my PhD and this dissertation possible (and even a little fun):

My favorite person in the world, Kris Liu. I know you're not much for compliments, so I will leave it at just one. Absolutely everything is better with you.

My longest supporters, my family. Being so far away has been difficult, but I love you all dearly, and I am so grateful for your endless encouragement.

My advisor, Steve Whittaker. You took a chance on me when you didn't have to, and my life will forever be better for having worked with you. Your compassionate mentoring has made me a more capable, confident researcher, and I will eagerly work together whenever possible in the future.

My intern mentors, Tara Matthews and Allison Woodruff. Tara, you turned what could have been a scary, abandoned summer at IBM into a wonderful new career direction. Allison, I am so appreciative for your guidance on this and other projects, and for your persistent advocacy on my behalf. I have thoroughly enjoyed working with both of you and am thrilled to be able to continue doing so!

My dissertation committee members, Doug Bonett and Sri Kurniawan. Your generous and thoughtful feedback improved this thesis immeasurably.

Finally, my first PhD advisor, Su-hua Wang. None of this would have been possible if you had not invited me to apply to work with you eight years ago, and I will always be grateful to you for giving me the start to my PhD.

Introduction

Large online companies increasingly affect people's day-to-day activities, choices, and experiences, and it is vital that we understand and measure the beliefs and attitudes that people hold towards these companies. Researchers in several fields have studied brand trust, the attitudes and beliefs people hold about a company's competence, integrity, and benevolence, as these beliefs reliably predict consumers' expectations and satisfaction when interacting with a company (Benedicktus, Brady, Darke, & Voorhees, 2010; Ha, 2004; Janiszewski & Meyvis, 2001; P. Lowry, Vance, Moody, Beckman, & Read, 2008).

However, recent research has shown that in addition to trust, people separately evaluate distrust as a parallel, distinct construct encompassing the skeptical, vigilant, and avoidant attitudes and beliefs that people hold towards companies (Benamati, Serva, & Fuller, 2010; Chang & Fang, 2013; Komiak & Benbasat, 2008; Moody, Galletta, & Lowry, 2014). Companies and researchers have been long interested in measuring and changing trust, but the presence of distrust as a parallel construct brings new questions to the forefront. When does distrust lead to better outcomes than trust? Do trust and distrust change in lockstep? Do equivalent changes in trust and distrust lead to different consumer outcomes? Each of these questions exposes a methodological gap: there are several valid, deployable scales to measure trust across a wide variety of companies, but there are, to date, no scales that are capable of measuring distrust across different companies.

The distrust scales that have been created thus far have been company- or industry-specific (Cho, 2006; D. H. McKnight & Choudhury, 2006; D. H. McKnight, Kacmar, & Choudhury, 2004), or focused only on dispositional aspects of distrust (Kehr, Kowatsch, Wentzel, & Fleisch, 2015; Kleiman, Sher, Elster, & Mayo, 2015). Companies and researchers must modify them substantially before deployment and cannot know whether those scales will perform adequately for a specific company. The goal of the current research is to address this gap by developing a novel scale, the Consumer Distrust in Specific Company Scale (CDISCS), which could be deployed across populations and companies by simply changing the name of the company being assessed. This thesis is structured as follows: first, I[1] review the core literature that informed the conceptualization and development of the proposed distrust scale. I then describe the three studies that comprised the creation and administrations to date of this scale. Finally, I share and interpret data on the current reliability and validity of the scale, describing the strengths of the scale as it stands as well as aspects of the scale that warrant additional research.

Literature Review

Trust appears to be driven by the amount of information an entity shares about itself, third-party verification of that information, and attributing competence to the entity (Afzal, Khan, Rehman, Ali, & Wajahat, 2009; Benedicktus et al., 2010; Ha,

[1]This dissertation research benefited greatly from the consultation and assistance of my advisor, Steve Whittaker. For clarity, I primarily use the singular "I" in this dissertation rather than "we", but I acknowledge and thank Professor Whittaker for his essential contributions to this research.

2004; P. Lowry et al., 2008). In contrast, distrust hinges more on an awareness of uncertainty in what to expect from an entity, and on the evaluation of whether expectations of the entity are or are not met over time (Chang & Fang, 2013; Cho, 2006; Komiak & Benbasat, 2008; D. H. McKnight & Chervany, 2001). Relying on information from third parties is crucial to both processes in different ways. Trust is enhanced by third parties as verifiers of information the company itself supplies, but distrust relies on third parties to detail possible negative outcomes that might result from encounters with the company.

The different processes underlying trust and distrust suggest they are not merely opposing ends of a single continuous scale. They are qualitatively different constructs. This is not only a theoretical point, as this distinction is important for researchers or advocates who wish to create interventions that change people's behavior, for instance to encourage people to improve their online privacy and security behavior, a critical task given pervasive threats to privacy and security online. Trust tends to encourage behavioral stability by continuing existing relationships, but distrust provokes people to reconsider their goals and how to achieve them (Lewicki, McAllister, & Bies, 1998). Thus interventions based on distrust might be more successful in changing consumer behavior than interventions aimed at changing trust. To evaluate this hypothesis, we need measures that can detect changes in distrust over time.

Trust and Distrust in Online Settings

Because trust has been more commonly studied, I first define trust and use it as a base to contextualize its cousin construct, distrust. Trust is an implicit belief that another entity will fulfill an expectation in the future (Baier, 1986; Berg, Dickhaut, & McCabe, 1995; Lewicki et al., 1998; Pettit, 2008). Trust underlies a wide spectrum of social and economic behaviors, because we depend on our assessment of others' future behavior when we consider what we will do in the present. As with most human behavior, trust is complex, and experts continue to argue about its exact definition. In order to ground this section, I begin with Baier's (1986) minimalist definition of trust: "A trusts B with C", meaning that agent A expects that agent B has sufficient ability, motivation, and integrity to satisfactorily handle a responsibility delegated to them to deal with valued thing C. As described by Baier (1986) and Pettit (2008), trust is a special case of reliance. To rely on an entity is to depend on the predictability of its habits. Reliance does not require consideration of the other party's good will; it only requires an implicit belief that another entity will continue acting the way it has in the past. For this reason, evaluating trust for an entity is more common at certain phases of relationships, for example when deciding whether to depend on an entity for the first time, or after the (un)successful completion of a trusted task.

I now turn to *distrust*, the degree to which one is wary, skeptical, or avoidant of another entity due to attributions of incompetence, malevolence, and/or dishonesty

(Lewicki et al., 1998; D. H. McKnight & Chervany, 2001). The definition of distrust has changed substantially over the past few decades. Researchers have at various times considered distrust to be an indicator of psychological disorder (Erikson, 1963), the absence of stable social ties (Lewicki et al., 1998), or simply a word mapped to the low end of trust, as seen in many trust scales (e.g., Sirdeshmukh, Singh, & Sabol, 2002). Perspectives on distrust began to shift in the late 1990s, in large part due to Lewicki and colleagues' (1998) paper on the relationship between trust and distrust. They argue that although relationships usually reward some degree of trust, interpersonal behavior is complex and often unpredictable. Overly high trust can lead to being taken advantage of, but there are examples of high trust relationships where the participants nonetheless maintain clear boundaries, a sign that some distrust is present in the relationship. For instance, spouses might share a bank account (evidencing high trust) but refuse to provide their email passwords to each other (evidencing some level of distrust). This would be difficult to conceptualize if distrust is only the opposite of trust, as high trust would presumably make the presence of distrust impossible. Rather than believing distrust to be an enemy of trust, or as a low state of trust, it should be considered as a responsible hedge against uncertainty, and as a distinct but parallel evaluation with trust. Trust may be responsible for increasing the benefits of mutual relationships , but distrust protects against being taken advantage of when trust should not be absolute within an interdependent relationship (Lewicki et al., 1998). Research since then has elaborated on definitions of distrust in

several ways, but theoretical and empirical studies have supported this conceptualization of distrust as separate from trust.

In this thesis, I focus on distrust for a specific class of entity: companies that an individual interacts with online. Research on online trust and distrust for companies and brands is common, but there is a clear imbalance between the availability and quality of measures that assess trust as opposed to distrust. The studies that have looked at distrust for online companies have found important distinctions between the processes and behaviors associated with distrust and trust. The situational features that consumers evaluate differ between trust and distrust (Komiak & Benbasat, 2008), and distrust has a greater effect on consumers' self-disclosure online than trust (Cho, 2006).

Distrust measures are far less common and often consist of trust items whose wording has been inverted, e.g., Trust: "This e-vendor will not engage in any kinds of exploitive and damaging behavior to customers", Distrust: "This e-vendor will engage in damaging and harmful behavior to customers to pursue its own interest" from (Cho, 2006). This method of distrust scale development makes sense given historical treatment of distrust as being identical to low trust. However, distrust is not the mirror image of trust, and therefore surveys assessing distrust should not be mirror images of trust scales. This was a key consideration during the creation of the current survey, requiring a greater understanding of how distrust differs from trust, particularly in the context of assessing distrust for a company. I next describe the

literature that most informed the construct definition and item creation phase of this thesis.

Prior Research on Distrust for Companies

When deciding whether to transact with a company, people make some level of judgment about their trust and distrust in that company. I draw most heavily on three models that define and explain the distrust processes people engage in when judging whether to choose or avoid a company: McKnight and Chervany's (2001) conceptual definitions of trust and distrust, Komiak and Benbasat's (2008) two-process view of trust and distrust building in electronic commerce, and Lankton and McKnight's (2011) model of mixed interpersonal-technology trust for online platforms like Facebook. Together, these models provided the basis for defining and operationalizing aspects of distrust of companies in the current survey.

McKnight and Chervany (2001) provide an interdisciplinary definition of distrust for companies that includes psychological, sociological, and economic aspects, which has been largely supported by empirical research (D. H. McKnight & Choudhury, 2006; H. McKnight, Kacmar, & Choudhury, 2003). This definitional model heavily influenced the focus of the current scale on aspects of distrust that influence the decision-making process rather than as a purely attitudinal construct. They draw heavily on Lewicki et al.'s (1998) conceptualization of distrust as an inherent tension of a relationship rather than as the opposite of trust. From this perspective, distrust is adaptive in that it both prevents an individual from being taken

advantage of due to misplaced trust, and is a natural reaction to violations of expectations in relationships. McKnight and Chervany build on this view by incorporating Luhmann's (1979) perspective that where trust reduces the complexity of social systems by reinforcing positive relationships, distrust functions as an impetus to develop new strategies such as reducing one's needs, seeking out new ways to fulfill them, or creating backup plans to satisfy unmet needs. In this view, low trust can differ from high distrust; going to a retail website that one has never seen nor heard of would likely evoke only low trust due to the lack of reputational information available; by comparison, if the web browser warned that the website has been reporting for spreading malicious files, that expands the space of possible negative outcomes from using that site, which would evoke distrust rather than only the absence of trust. McKnight and Chervany's (2001) definition of distrust contains

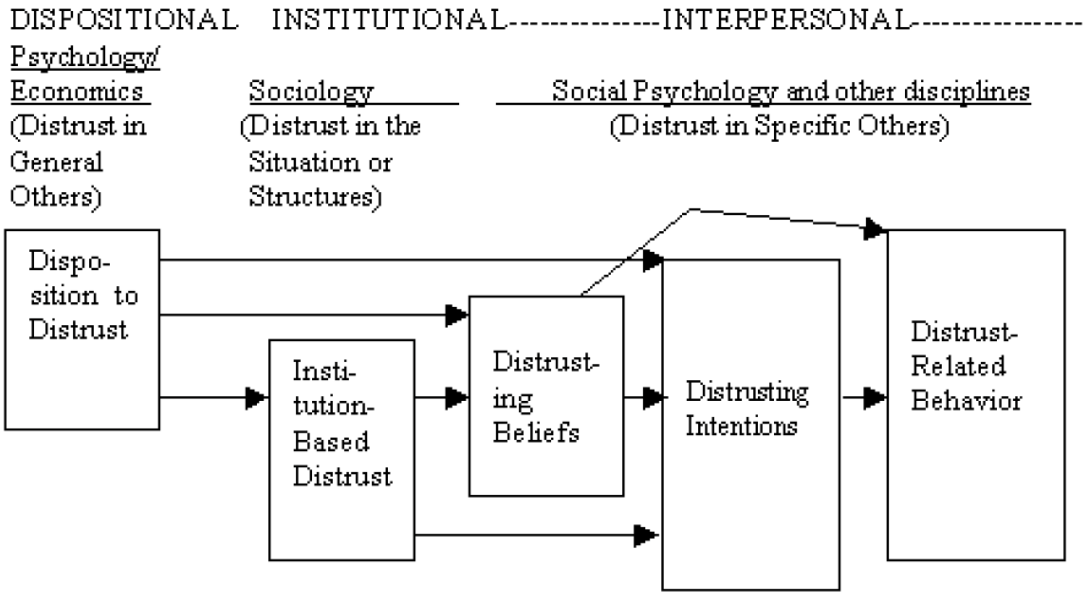

Figure 1. Interdisciplinary constructs in McKnight and Chervany's definitional model of distrust (McKnight & Chervany, 2001).

dispositional, institutional, and inter-entity aspects (see Figure 1), which I will describe in greater detail later in this section.

Komiak and Benbasat (2008), by contrast, provide a more contextualized, process-oriented view of how trust and distrust develop over repeated interactions with a branded service, i.e., a recommender system from ActiveBuyersGuide.com. They traced patterns of trust and distrust over longitudinal interactions with the recommender system by qualitatively analyzing participants' think-aloud sessions as they used the service. Nine processes related to trust and distrust emerged in that analysis, clustering around knowledge, usability, successful outcomes, and attributions. It is critical to note that several of these differed substantially between trust and distrust, supporting the view that these are different processes (see Figure 2).

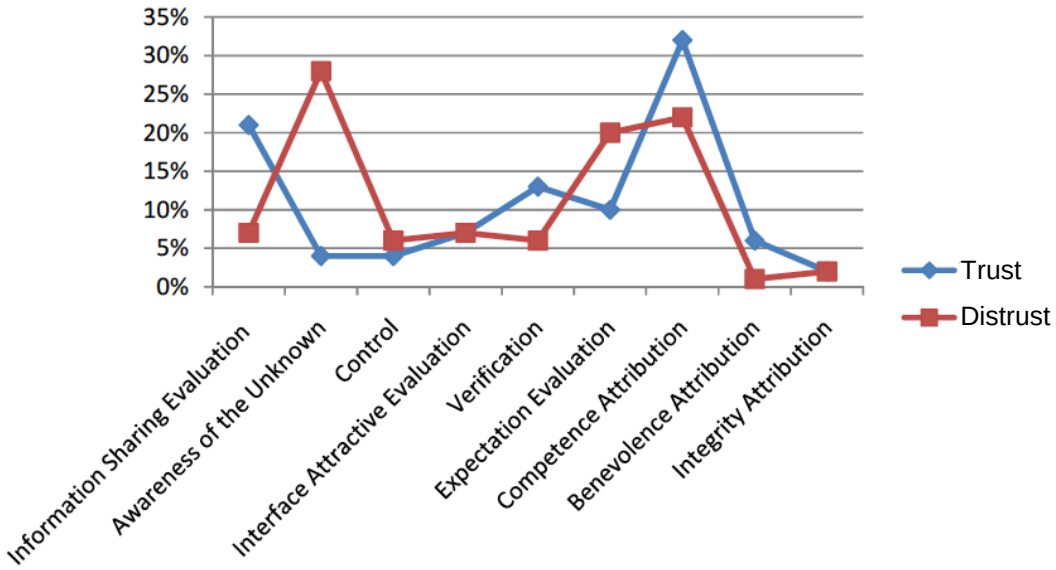

Figure 2. Differences between trust and distrust processes for a branded recommender system (Figure from Komiak & Benbasat, 2008).

Trust processes tend to be driven by how much information an entity makes available about itself, and third-party verification of that information; by comparison, distrust depends more on awareness of what might go wrong, and the extent to which the entity fails to fulfill its commitments (Komiak & Benbasat, 2008). The processes they identified as being most common in distrust evaluations (e.g., Awareness of the Unknown, Expectation Evaluation, Competence Attribution), were used during the construct definition and item creation phase of the current scale, enabling us to develop a scale that was distrust-specific, not simply a negatively-worded trust scale.

I also draw on Lankton & McKnight's (2011) study showing that trust beliefs for online platforms manifest as a combination of interpersonal and technology evaluations. People evaluate trust for individuals in terms of (1) competence, (2) benevolence, and (3) integrity. Technology is evaluated in terms of the respectively analogous constructs (a) functionality, (b) reliability, and (c) helpfulness. In their study, Lankton and McKnight tested whether Facebook users' trust for Facebook is best modeled using the interpersonal trust constructs (1), (2), and (3); using the technology trust constructs (a), (b), and (c); or as a hybrid of the two, creating second order factors that combine each pair of analogous constructs as (1+a), (2+b), (3+c). They found the hybrid approach outperformed either of the other approaches, indicating that when people evaluate trust for an entity like Facebook, they consider it as both an entity and as a technology. I extend their approach to distrust, including

distrustful belief items mapping to each of these second-order factors during the item creation process described in Study 1 below.

Because I adapted aspects of all three models in the present study, I organize the remainder of this section around the constructs themselves rather than the models. For each construct, I first provide a definition drawn from these models, then review empirical findings relating to the construct, and finally describe my implementation of the construct for the current survey.

Disposition to distrust. The disposition to distrust represents an individual's general propensity to distrust others across situations (D. H. McKnight & Chervany, 2001). Personality variables such as this are common components of theories that explain responses to potential threat situations, such as protection motivation theory (Floyd, Prentice-Dunn, & Rogers, 2000) and privacy calculus models (Li, 2012), as they explain individual differences in the tendency to interpret ambiguous risk in environmental stimuli. Several psycho-diagnostic scales contain distrust or mistrust items which I drew upon when constructing my subscale on this construct, including the Mistrust subscale in the CAT-PD (Simms et al., 2011), the Vigilance factor in the 16 Personality Factor Questionnaire (Cattell & Mead, 2008), and Tellegen's Multidimensional Personality Questionnaire (Cain, Lukowitsky, & Wright, 2014).

McKnight and Chervany divide this disposition into two main subconstructs: *suspicion of humanity*, the assumption that others are usually dishonest, malevolent, incompetent, or unpredictable; and *distrusting stance*, the assumption that regardless

of suspicion, acting as though others will let you down leads to better outcomes (D. H. McKnight & Chervany, 2001). In more recent work, McKnight has shown that the disposition to distrust is a better predictor of how users interpret high-risk websites than the disposition to trust (McKnight, Kacmar, & Choudhury, 2004). Kleiman and colleagues similarly found that a more distrustful disposition leads to greater consideration of alternatives when deciding how to achieve a goal (Kleiman et al., 2015). In the current study, I use McKnight et al,'s (2001) definition of a disposition to distrust, creating items that assess *suspicion of humanity* and *distrusting stance* to account for how personality differences affect individual's interpretations of companies they interact with.

Institution-based distrust. I now turn to Institutional-based Distrust, the aspects of a company that affect a person's evaluation of distrust for that company. McKnight and Chervany (2001) define this construct as the lack of conditions that are conducive to successfully navigating a risky situation. They divide this into a *lack of structural assurance* due to missing safeguards like contracts of guarantees; and a *lack of situational normality*, which they leave largely undefined but presumably is meant to capture ways that a company's practices differ from norms set by competitors or established corporate standards. This definition implies that institution-based trust is the norm, with distrust of institutions being purely about what is missing, but other work emphasizes situational features that, when present, do increase distrust. Komiak and Benbasat (2008) found that customers' distrust

processes frequently include evaluation of knowledge and experience they build when interacting with a brand. Participants' uncomfortable experiences such as poor recommendations or other violated expectations led to suspicion, discomfort, and distrust. This suggests that McKnight and Chervany's model could benefit from inclusion of negative experiential indicators of institution-based distrust rather than merely the absence of positive indicators, which may evidence low trust but not necessarily high distrust. Several studies have shown that trust for a brand increases with the number of positive interactions, i.e., problems successfully solved (Afzal et al., 2009; Benedicktus et al., 2010). It is conceivable that repeated negative interactions might have a similarly strong effect on distrust.

Beyond prior experience which Komiak and Benbasat studied, another potential indicator of institution-based distrust is word-of-mouth. Several studies have shown that word-of-mouth highly affects perceptions of brand reputation. Positive reviews are associated with higher trust for products, including for books (Chevalier & Mayzlin, 2006), hotels (Sparks & Browning, 2011), and, interestingly, for reviews themselves (Chen, Dhanasobhon, & Smith, 2001). My pilot interviews also emphasized that *negative* reviews, not lack of structural assurance, play a large role in developing distrustful beliefs about companies. However, empirical work looking at how negative reviews affect distrust for companies remains rare. To address this gap, I include word-of-mouth as an aspect of institution-based distrust in the current survey to facilitate future research in this area.

Distrustful beliefs. Distrustful beliefs are present and similar in each of the distrust belief models I use here, as they comprise what is typically considered to be the core evaluation of distrust itself: attributions of an institution's competence, benevolence, and integrity (Komiak & Benbasat, 2008; Lankton & McKnight, 2011; D. H. McKnight & Chervany, 2001). Here, competence refers to the entity's ability to fulfill an expectation, benevolence refers to the entity's motivation to help or hinder, and integrity refers to the (dis)honesty with which an entity acts when fulfilling commitments. As previously mentioned, Lankton and McKnight (2011) also include technology-related trust components here, such that functionality, helpfulness, and reliability are related to competence, benevolence, and integrity, respectively.

Several distrust beliefs scales have been created, though they often have features that preclude incorporating items directly; some are only appropriate for evaluating a specific industry or website (Rose, Peters, Shea, & Armstrong, 2004; Shea et al., 2008), whereas others have clarity issues due to item and response option wording (e.g., "I am uneasy about whether [Entity] is sincere and genuine. [strongly disagree-strongly agree]") (P. B. Lowry, Schuetzler, Giboney, & Gregory, 2015; D. H. McKnight & Choudhury, 2006). Like Institution-based Distrust, theory suggests that these constructs should focus on negative indicators of these constructs rather than merely the absence of positive indicators. Therefore, I created several items in the current survey that measure, for instance, attribution of malevolence rather than a lack of benevolence.

Distrustful intentions. Fairly straightforwardly, this construct refers to an individual's intention to avoid an entity in the future, even if it might lead to negative consequences to the individual themselves (D. H. McKnight & Chervany, 2001). This is a self-rating of how likely and how confident an individual is that they would go out of their way to avoid relying on an entity in the future. As a prospective, hypothetical evaluation, it should be noted that people are notoriously unreliable at estimating their likelihood of sticking to such behavior e.g., the privacy paradox in which even people who claim strict beliefs about responsible privacy behavior continue to make privacy-harming decisions (Acquisti, Brandimarte, & Loewenstein, 2015; Acquisti & Grossklags, 2004; Norberg, Horne, & Horne, 2007). In McKnight and Chervany's model, they propose that distrustful intentions should be a moderating variable through which disposition, institutional features, and beliefs affect behavior. As a result, I include Distrustful Intentions as a construct in my proposed scale.

Distrustful behavior. Distrustful Behavior is an important correlate to distrustful beliefs and intentions. Distrustful Behavior comprises actions that an individual takes to strategically reduce vulnerability from another entity (D. H. McKnight & Chervany, 2001). This gets at the heart of what distrust is hypothesized to affect, the likelihood of taking vigilant actions that hedge against harm (Chang & Fang, 2013; Moody et al., 2014). There are several strategies that people can engage in with companies to reduce vulnerability, such as seeking out information in reviews

or policies, providing false information to avoid privacy breaches, maintaining redundant options or back-ups in case of service outages or product breakage, or switching to competitors when alternatives are available. Although distrustful beliefs and intentions encompass the cognitive and metacognitive aspects of distrust, Distrustful Behavior provides evidence that distrust affects the way one accomplishes one's goals and is therefore the strongest indicator of distrust. Other scales that focused on a single industry or company have included items or subscales assessing behavior that indicate distrust in those specific settings, e.g., "How likely are you to

Creation	Construct Definition	Study 1
	Item Creation	Study 1
	Cognitive Pretesting	Study 1
Reliability	Retest (2 weeks between)	Study 1
	Differential across demographics	Study 2
	Internal Consistency	Study 1, 2
Validity	Face	Study 1, 2, 3
	Content	Study 1, 2
	Convergent	Study 2
	Discriminant	Study 2
	Criterion	Study 3

Table 1. Summary of the survey creation and validation process described in this thesis.

do most of your future travel on this airline?" (Sirdeshmukh et al., 2002). Because the goal of the current research is to create a scale that is flexible across companies, and because distrustful behavior varies so heavily across products, services, and companies, I do not include it as a subscale in this survey.

I build on work that has defined distrust as a parallel but distinct process from trust (Komiak & Benbasat, 2008; Lewicki et al., 1998; D. H. McKnight & Chervany, 2001; D. H. McKnight & Choudhury, 2006; Saunders, Dietz, & Thornhill, 2014; Schul & Peri, 2015), which has led me to create a novel, theoretically driven questionnaire that assesses individuals' distrust in a company. The scale I developed in this thesis assesses features of distrust that include the disposition of the respondent to distrust, institutional features of distrust, as well as distrustful beliefs and intentions held about the company. The scale creation and validation steps I conducted in this thesis are summarized in Table 1. In Study 1, I describe the theory-led creation, piloting, and reliability testing of the proposed scale. In Study 2, I show the results of a full administration of the scale, including assessment of the scale's current reliability, validity, and adherence to theoretical models. Finally, Study 3 demonstrates the criterion validity of the scale by an intervention study where participants' distrust was assessed before and after learning about a real-world scandal with their actual bank.

Study 1: Scale Creation and Pilot Administration

Creating the initial pool of items for the CDISCS required integration of the theoretical and methodological resources described in the prior section to ensure the relevant distrust constructs were adequately represented in the scale. Study 1, described in this section, includes the construct definition, item creation, and cognitive pretesting that led to the first version of each of the four subscales considered for inclusion in the scale. These initial subscale versions were then piloted with participants recruited using Mechanical Turk to assess their performance, including distributional characteristics, retest reliability, and internal consistency.

Method

Design. This survey creation process followed the Tailored Design Method approach (Dillman, Smyth, & Christian, 2014). In the Tailored Design Method, external feedback is sought early and often, including through expert consultation, cognitive pretesting, verbal probes, and piloting, in order to reduce four key types of error in the survey: *sampling* (who is contacted to complete the survey), *coverage* (the difference between who is sampled and who completes the survey), *nonresponse* (differences between the participants who do and do not answer a particular question), and *measurement* (the deviation from the participant's true state and their response to a particular item). I address each of these sources of error in the design of this survey and point out how these challenges were tackled for the current survey.

In the last twenty years, survey development has largely transitioned from live telephone interviews and blind-mailed surveys to self-administered online deployments of questionnaires (Callegaro, Manfreda, & Vehovar, 2015; Dillman et al., 2014). Compared to pen-and-paper, online questionnaires tend to have similar or higher response rates and reliability, and similar or lower items nonresponse rates (Dillman et al., 2014; Greenlaw & Brown-Welty, 2009; Ritter, Lorig, Laurent, & Matthews, 2004). Online surveys are drastically cheaper and faster to obtain results compared to pen-and-paper surveys (Greenlaw & Brown-Welty, 2009), and they have several features that are difficult or impossible to include in older modalities. For example, online surveys allow the survey designer to automatically pass a respondent's answer to an early question into later items, reducing measurement error that might result if the respondent was required to remember their answer to previous items during the rest of the survey (Callegaro et al., 2015). Online surveys also facilitate cost-effective deployment of a survey to non-student respondents, reducing sampling error by not limiting to undergraduate participants.

A particularly well-studied method of recruiting survey respondents online is to use crowdsourcing platforms such as Amazon's Mechanical Turk, a website where workers can complete tasks for a requester in return for compensation (Buhrmester, Kwang, & Gosling, 2011; Mason & Suri, 2012). Although its population includes a greater proportion of young, male, and White/European-American users compared to the general US population, several studies have shown that recruiting through

Mechanical Turk can achieve similar results to telephone or other methods of data collection, particularly for self-administered surveys (Buhrmester et al., 2011; Holden, Dennie, & Hicks, 2013; Rouse, 2015; Simons & Chabris, 2012). Survey validation is increasingly conducted through recruitment on Mechanical Turk (Egelman & Peer, 2015; Holden et al., 2013), due to its high speed, low cost, and ease of integration with secure survey platforms such as Qualtrics that allow for demographic filtering (Peer, Paolacci, Chandler, & Mueller, 2012).

Construct definition. As noted above, trust has been well-studied in psychological research and across various fields for decades, however distrust has received much less attention. I follow the definitions of distrust drawn primarily from McKnight & Chervany (2001), and Lewicki and colleagues (1998), who both created taxonomies of distrust features that I combined to operationalize constructs for this survey. The three features of distrust that I include as originally conceived by McKnight and Chervany are the Disposition to Distrust, Distrustful Beliefs, and Intentions to Distrust. I briefly redescribe these, introduce my final subscale, a modified version of McKnight and Chervany's (2001) Institutional Features construct in response to feedback I obtained from expert interviews.

First, Disposition to Distrust refers to a person's stable tendency to avoid depending on others across contexts. It is comprised of two main subconstructs: suspicion of humanity, the degree to which one assumes that the general population is deserving of distrust; and distrusting stance, the belief that one achieves better

outcomes across contexts by behaving as though others are malevolent or otherwise unreliable.

Distrustful Beliefs are a person's beliefs about the degree to which a particular entity manifests each of three key aspects that provoke distrust: incompetence, only being capable of doing poor quality work; malevolence, the motivation to hurt rather than help; and dishonesty, the tendency to lie rather than tell the truth and to break rather than keep commitments.

Intention to Distrust refers to an individual's subjective perception of their likelihood to avoid depending on an entity in the future.

Finally, my Institutional Features of Distrust subscale differs substantially from McKnight's Institution-based Distrust construct. This different approach followed from pilot interviews conducted to conceptualize the role distrust plays in people's evaluations of online companies. McKnight focuses on lack of structural assurance and lack of situational normality as the main aspects assessed when evaluating distrust for an entity, but those concepts were largely absent in the 12 expert interviews I conducted to determine the main ways that interviewees with extensive online experience assessed their distrust for a company. Instead, interviewees mentioned word-of-mouth and prior experience as the ways that they evaluated to what degree they distrust a company. Consistent with this, prior research on brand trust has also shown word of mouth and prior experience to be important predictors of consumer behavior (Awad & Ragowsky, 2008). And process models of

distrust emphasize that unmet expectations lead differentially more to distrust than trust development (Komiak & Benbasat, 2008). I therefore used these subconstructs in lieu of McKnight's.

Item creation. Countless resources exist for helping create good survey items, providing useful rules of thumb such as how to reduce measurement error by building questions around one construct each, simplify wording, and ensuring that later items are not dependent upon earlier items (Callegaro et al., 2015; Dillman et al., 2014, 2014; Marsden & Wright, 2010). In addition to these general texts, I drew on empirical work to improve several aspects of the survey, such as to consider the effects of item characteristics like negatively-keyed items (e.g., "How sure are you that you would [*not choose*/*avoid*] [Company]?") on the eventual factor structure of the scale (Schmitt & Stults, 1985; Spector, Van Katwyk, Brannick, & Chen, 1997; Woods, 2006). Following Yeager and Krosnick's advice (Yeager & Krosnick, 2012), I also designed items to not depend on prefacing statements about "some people" and "other people".

I also relied heavily on resources for choosing and designing response scales. First, I used Pasek and Krosnick's (Pasek & Krosnick, 2010) guidelines to employ five-item scales for unipolar concepts (e.g., "How corrupt is [Company]"), and seven-item scales for bipolar concepts (e.g., "How good or bad is [Company] at what it does?). I minimized my use of agree-disagree scales in favor of bespoke response options for each question to reduce boredom and prevent acquiescence response bias

(Knowles & Nathan, 1997). After a suggestion during a cognitive interview, I referred to (Rohrmann, 2007) to choose response options that were shown in that study to be judged by participants as nearly equidistant from each other from a 0-10 point scale, i.e., “not at all” = 0/10, “slightly” = 2.5/10, “moderately” = 5/10, “very” = 7.9/10, “extremely” = 9.6/10.

To develop the individual items for each subscale, I drew on existing trust and distrust scales (Kehr et al., 2015; Kleiman et al., 2015; D. H. McKnight & Choudhury, 2006; Montague, Kleiner, & Winchester, 2009; Shea et al., 2008)(McKnight & Choudhury, IPIP, Yamagishi & Yamagishi, Montague et al., Huynh & Hirschheim), with the provisos mentioned above. No items were incorporated directly from prior scales, but I note where my items were adapted or derived from existing items. The decision not to import items from other scales was due to both content and structural concerns. McKnight and colleagues’ distrust questionnaire items revolve heavily on agreement-based response scales to uncertainty-based framing. While distrust is assuredly related to uncertainty, such framing limits the types of distrust that these items measure. For instance, McKnight et al.’s (D. H. McKnight & Choudhury, 2006) items like “I feel worried that...” or “I feel uncertain about” do not leave room for a participant to express that they are sure about a negative outcome with a company. By contrast, items in the current survey were worded to allow respondents to rate specific aspects of their distrust for the company across various dimensions, including the likelihood of misbehavior, the

strength of their conviction about the company's dishonesty, and the degree of mismatch between their expectations and experiences with the company.

Another key design consideration when creating items was that they apply across different types of companies. To keep that in mind during item creation, I created a table containing exemplar companies who provide different types of products and/or services, including financial, utility, retail, online brands (e.g., Google, Facebook), etc. to ensure that items would apply well to each type of company. The companies in this table were drawn from Alexa.com's list of the most used company websites ("Top Sites in United States - Alexa," n.d.), as well as by searching for companies with varying reputations using Consumer Reports-type resources, e.g., Consumerist.com ("Consumerist," n.d.). This requirement that items apply to companies in different fields limited the pool of items I could draw from previous scales, as those that have assessed distrust directly have tended to be either website-specific (D. H. McKnight et al., 2004) or applicable to only specific industries such as retailers (Cho, 2006). Nevertheless, I was able to adapt some of these items for use in the current scale, e.g., McKnight's "It is uncertain whether LegalAdvice.com would keep its commitments" was adapted to create a new item: "I believe that [Company] (never...always) keeps its commitments".

Items were created to fit the 4 subconstructs identified in the prior section, so that the topic space for each construct was mapped fully. For example, in the distrustful beliefs section, I created 7 items for incompetence (e.g., "[Company]

(never...always) does a bad job), 15 items for malevolence (e.g., ([Company]'s values are (not at all...completely) different from my values)), 11 items for dishonest misbehavior (e.g., "If they make a promise, [Company] (never...always) breaks it), and 4 items for unpredictability (e.g., "Relying on [Company] means not knowing when things will go wrong (Completely disagree...Completely agree)). During pretesting and piloting, items were frequently revised, and some items were added to or removed from subscales, based on criteria I detail below.

I supplemented items that were adapted from existing scales with newly created items derived from by studies mining natural descriptions of distrustful situations. One useful resource for this was Montague et al., (2009) which identified words and phrases used to describe medical technology that participants trusted or distrusted to different degrees. This type of naturalistic resource is valuable given the goal to incorporate inter-entity and technology-based distrust concepts in the survey based on (Lankton & McKnight, 2011). Because these descriptions are in language that people naturally use rather than being researcher-driven, measurement error should also be reduced. Additionally, these natural descriptions allowed me to vary the strength of subscale items systematically. For example, Montague et al. found that the word "deceitful" was a higher indicator of distrust than the word "uncaring". Despite both being unflattering words one could use to describe a company, "deceitful" is a more stringent indicator of distrust than "uncaring".

Unlike the other subscales, the Distrustful Intentions items were initially created as Guttman scales: sets of 4 binary yes/no items that have an essential order to them (Stouffer et al., 1950). Pilot participants disagreed about the proper ordering of the items, suggesting that approach might be a source of measurement error if deployed widely. In response to that feedback, the Guttman scales were adapted into Likert-style items similar to those in the other subscales.

Cognitive pretesting. I relied on several resources to design cognitive pretesting that would be most useful given the characteristics of the desired survey. Willis (Willis, 2004), Sebastiani et al. (Sebastiani, Tinto, Battisti, & De Palma, 2013), and Dillman et al. (Dillman et al., 2014) all recommend the use of 1:1, in-person cognitive interviews to pretest novel questionnaires prior to a pilot administration. The two recommended options for these interviews are think-aloud, in which each interviewee self-narrates to the interviewer as they complete the entire survey, and verbal probing (Sebastiani et al., 2013; Willis, 2004), where the interviewer strategically asks questions about specific aspects of individual questions as they are answered. Although think-aloud is a good choice for establishing the process by which verbose participants answer questions, the self-narration is difficult for some interviewees, poorly resembles the circumstances under which pilot participants will typically answer each question, and may change participants' thought processes by requiring them to multitask rather than focus on the item at hand (Ericsson & Simon, 1998). Verbal probes, on the other hand, allow for a more natural questionnaire

completion and greater interviewer control over what specific questions are asked, at the expense of potential bias if the interviewer asks leading questions. I chose to use verbal probes, using the concurrent method described by Willis (Willis, 2004): (1) the interviewee reads and answers the question without interviewer interference, (2) the interviewer asks a probe question, (3) the interviewee answers the probe, with steps 2 and 3 repeating for each additional probe on that question. This reduces bias that would be introduced if the interviewer asked a probe question before the interviewee responded to the original item on their own. In cases where the interviewee asks a clarifying question before answering, the interviewer asks the respondent to respond to the item first, and that they will answer the interviewee's question afterwards. Some probes were scripted and used systematically for the same question to all interviewees, but many were spontaneous based on the interviewee's prior answers to questions or probes.

Each subscale went through specific cognitive pretesting, following the approach described above. This pretesting involved 1:1 administrations of that questionnaire section to expert and non-expert participants who provided formative feedback on item wording, visual design, response scales, and construct thoroughness. In addition, I conducted verbal probe-based cognitive interviews for the Distrustful Beliefs subscale, as it contained the most novel items. Each cognitive interview was structured so that the cognitive interviewee took the survey using

Qualtrics, using the identical set-up that participants would later see during the pilot administration.

Each subscale saw alterations due to interview feedback. Items were created, reworded, removed, or altered to increase clarity, reduce confusion, and measure more aspects of each construct. I also made changes to the survey design as a result of the pretesting feedback, including the addition of an open-ended question to contextualize and provide convergent validity information for a respondent's answers to the survey: "Please tell us, in your own words, what you think about [Company] as a company, and how you came to feel that way". This item was used both to check whether participants were paying attention during the survey, as well as to generate new items based on how respondents describe their experiences with companies they distrust.

Scales used in piloting. The revised scales used in the pilot administration were comprised as follows: disposition to distrust (23 items), institutional features of distrust (21 items), distrustful beliefs (40 items), distrustful intentions (8 items). The number of items in each subscale differs based on characteristics of those subscales: the disposition and institutional features subscales have two subconstructs, whereas the larger beliefs subscale has three constructs, and the smaller intentions subscale is formed to assess a single construct.

Pilot administration. Following Egelman and Peer's validation of their Security Behavior Intentions Scale (Egelman & Peer, 2015), I used Amazon

Mechanical Turk to recruit participants for piloting each subscale. In order to maintain participants' privacy in accordance with the IRB's advice, I used Mechanical Turk only for recruiting and paying participants, directing them to complete the survey on Qualtrics, a secure survey platform approved for use with human subjects by the UC Santa Cruz Institutional Review Board ("Internet-Based Research," 2015). Qualtrics allows participants' responses to an early question to be inserted into later questions and response options. I used this in most questions to insert the name of the target company that the participant identified at the start of the survey, e.g., a participant who wrote in "Jobeline" as a company they have discontinued using would see "In the future, I intend to avoid [Company]" as "In the future, I intend to avoid Jobeline". This reduces nonresponse and measurement error by making it easier for participants to complete the survey without needing to mentally insert the name of the company in each question; it also reduces sampling error by not requiring that participants have used an experimenter-chosen company to qualify to take the survey. I also enabled "ballot box stuffing prevention" in Qualtrics, which disallows taking the same survey twice from the same IP address. This reduces sampling error by preventing any person from being represented twice in the same sample, e.g., if they wanted to take the survey twice to earn twice the fee from Mechanical Turk. In addition to Qualtrics, I used TurkPrime, a service that integrates with Mechanical Turk and Qualtrics to validate participants' completion of the survey with a dynamically-generated code, assigns qualifications to Mechanical Turk

workers that prevent them from taking different conditions of the same survey, and allows researchers to contact Mechanical Turk workers after completion of an assignment, which I used to contact and pay bonuses to participants in the 2-week retest. Each of the subscales was piloted separately. The flow of the survey administration was as follows (see Figures 3 and 4 below):

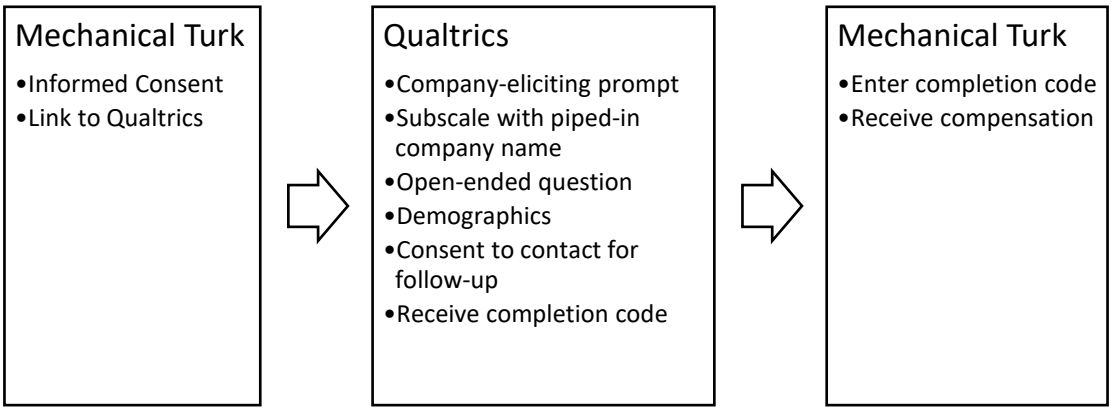

Figure 3. Procedure flow for test pilot administration of each subscale.

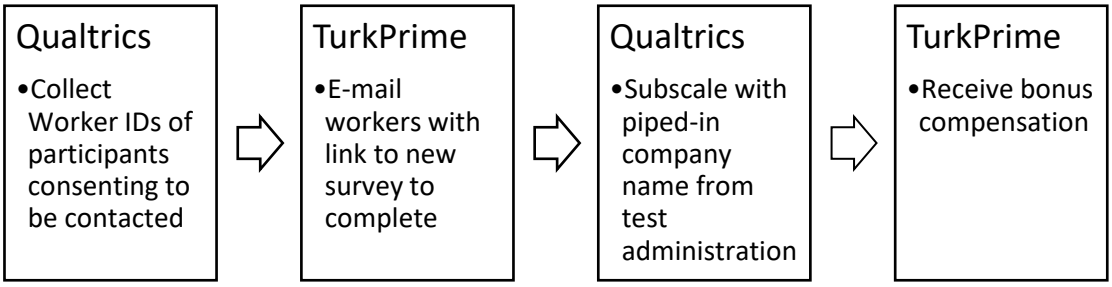

Figure 4. Procedure flow for retest pilot administration of each subscale.

The participant viewed the call for participation on Mechanical Turk, which offered $2 payment for an estimated 12-minute survey study about a company they have previously used, or for the Disposition subscale, about their social beliefs.

Participants read the informed consent form, and if they agreed to participate in the study, they were directed to click a link to the survey on Qualtrics. They inputted their Mechanical Turk worker ID and answered a company-eliciting prompt, a technique I drew from interpersonal trust survey validation studies that ask participants to rate a friend, a spouse, a boss, a childhood antagonist, etc. (Larzelere & Huston, 1980). Three such prompts were used, "Thinking about companies whose online products or services you have used in the last year, what is the name of [the company you have had the most positive experiences with/the company you have had the most negative experiences with/a company whose products, services, or website you have stopped using]? Please take your time in answering this question, as the remainder of the survey will be about the company you name here." The company they named was inserted in relevant places throughout the remainder of the survey.

In the distrust subscale section of the survey, participants saw each survey item, grouped by construct and presented with 5-7 items per page. In the subscale section, no questions required a response in order to continue the survey, in order to reduce nonresponse bias from participants who might leave the survey incomplete if forced to answer a question that makes them uncomfortable (Callegaro et al., 2015). Following the subscale items, participants answered a single open-ended question: "Please tell us, in your own words, what you think about [Company] as a company, and how you came to feel that way". The participant then completed the demographics section, which asked the participant's age, gender, educational

attainment, income, and state of residence. I asked the demographics towards the end to reduce the risk of participants bouncing off the survey (i.e., leaving without answering a single question) due to discomfort at being asked potentially sensitive questions at the outset (Callegaro et al., 2015; Dillman et al., 2014). Finally, participants were asked whether they agreed to be contacted through Mechanical Turk about the follow-up survey in the next couple of weeks. 100% of participants agreed to be contacted for the follow-up survey. After submitting the survey, they received a completion code that they entered on Mechanical Turk to submit the survey and receive compensation. One participant was unable to generate the completion code, so they contacted me by e-mail, and I manually verified their completion of the survey. Every participant who submitted the survey was paid $2 regardless of data quality. Participants whose responses were completely uniform, or "straight-line", were added to a block list for future Mechanical Turk studies but were paid for their time on the work they completed.

To conduct the retest administration, I collected the Worker ID and the company they responded about from the participants' responses. After verifying they agreed to be contacted, I created a new retest version of the original survey that the participant had taken. I automatically inserted into the retest version the company name that each participant had responded about during the first session. Two weeks after the initial pilot administration for each subscale, I messaged each Mechanical Turk worker who had participated, offering them a $2 bonus to their original payment

if they completed a follow-up survey. Coverage in the retest phase was high, >60% participation. The survey was identical to the prior version, except it lacked the open-ended and demographic questions. This was to avoid burdening participants and to reduce nonresponse error.

Pilot data analysis. The goal of Study 1 was to assess item and subscale performance to reduce the initial pool of items before administering the full scale to a larger sample. In this phase, items were selected based on a variety of features, including descriptive statistics, retest reliability, conceptual representativeness, internal consistency, and exploratory factor analysis. The distributions, central tendency, and variability of each item were assessed to select items with approximately normal distributions and without ceiling or floor effects. There are several options for measuring retest reliability for survey items, including the percentage of exact agreement for each item, the Pearson correlation coefficient, the intraclass correlation coefficient (ICC), and Cohen's kappa (Tabachnick & Fidell, 2007; Tavakol & Dennick, 2011; Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013). Each statistic has advantages and disadvantages, but for ease of interpretability, I report the Pearson's correlation coefficient for each item, and for each of the four proposed subscales.

I expect retest reliability to differ across the subscales. The Disposition subscale should have the highest retest reliability due to it eliciting a stable characteristic. By comparison, perceptions of a company's incompetence ought to

more mutable because of ongoing interactions with the company or competitors. Items with retest reliability of $r < .50$ will be considered for removal, and I will consider the subscales to be reliable if they should have retest reliability of $r > .70$. Internal consistency is a measure of interrelatedness between items in a scale, which I measured using exploratory factor analysis (EFA) as well as the standard Cronbach's alpha. To claim the subscales are internally consistent, I aim to show Cronbach's alpha greater than .70, although $\alpha > .90$ would be considered a sign that the subscale in question may be contain redundancies or be too unidimensional (Streiner, 2003). Items that load poorly onto a single factor EFA, or that reduce Cronbach's alpha substantially were considered for removal.

Results

The initial set of items were assessed for potential inclusion in the final scale using three metrics: (a) analysis of response distributions to ensure appropriate variability and relative statistical normality in responses, (b) assessment of retest reliability by correlating responses between the two deployments, and (c) assessing coherence of items in the hypothesized subscales by calculating internal consistency of each subscale. I now describe these analyses for each of the four proposed subscales based on McKnight and Chervany's (D. H. McKnight & Chervany, 2001) definitional model of distrust. Retest participation ranged from 48% to 71% for the four subscales.

Disposition to distrust. In total, 37 out of 50 participants completed both administrations of the Disposition subscale. The Pearson correlations between the first and second administrations for individual items ranged from .47 (95% CI [.17, .69]) to .88 (95% CI [.78, .94]), averaging .71 (*sd* = .12), indicating high retest reliability over two weeks, which is to be expected given that disposition towards distrust is hypothesized to be a stable personality trait.

Items were first removed due to low variability, bimodal distributions, and substantial skew. Next, items with low retest reliability ($r < .50$) were considered for removal. Finally, a factor analysis was conducted to determine whether any of the remaining items loaded poorly on the factor that most items tapped into. The items kept as the disposition subscale in the next phase were DIS2, DIS6, DIS9, DIS10, DIS16, and DIS17. The resulting subscale had high retest reliability and internal consistency: Pearson's *r* between the two administrations was .92 (95% CI [.86, .96]), and Cronbach's alpha was .92 for the six-item subscale.

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
DIS1	People are no good.	{1, Completely disagree}... {5, Completely agree}	3.42	1.13	.88	[.78, .94]
DIS2	**People are better off assuming that others will...**	**{1, Never let them down}... {5, Always let them down}**	**3.20**	**1.11**	**.87**	**[.76, .93]**
DIS3	I believe that most people...	{1, Never tell the whole truth}... {5, Always tell the whole truth}	2.96	1.01	.82	[.68, .90]
DIS4	People are safer if they assume the worst will happen.	{1, Completely disagree}... {5, Complete agree}	3.12	1.21	.82	[.68, .90]

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
DIS5	Assuming that other people will let you down is...	{1, An extremely bad strategy}... {7, An extremely good strategy}	4.40	1.80	.80	[.64, .89]
DIS6	**I assume people that I meet are...**	**{1, Not at all selfish}... {5, Extremely selfish}**	**2.82**	**1.06**	**.80**	**[.64, .89]**
DIS7	If you put your trust in other people, they will...	{1, Never let you down}... {5, Always let you down}	3.04	0.81	.78	[.61, .88]
DIS8	When someone offers to help me, I assume they have hidden motives for doing so.	{1, Completely disagree}... {5, Completely agree}	2.78	1.22	.78	[.61, .88]
DIS9	**In general, I am...**	**{1, Not at all suspicious of people I don't know well}... {5, Extremely suspicious of people I don't know well}**	**3.02**	**1.14**	**.76**	**[.58, .87]**
DIS10	**Most people...**	**{1, Never lie when they would benefit from doing so}... {5, Always lie when they would benefit from doing so}**	**3.12**	**0.90**	**.75**	**[.56, .86]**
DIS11	Until they prove me otherwise, I assume people are...	{1, Not at all unreliable}... {5, Extremely unreliable}	2.94	1.30	.75	[.56, .86]
DIS12	Being skeptical about other people's intentions is...	{1, Not at all useful}... {5, Extremely useful}	3.42	1.13	.74	[.55, .86]
DIS13	I find it hard to forgive others.	{1, Completely disagree}... {5, Completely agree}	2.86	1.29	.74	[.55, .86]
DIS14	People try to get away with as much as they can.	{1, Completely disagree}... {5, Completely agree}	3.46	1.05	.74	[.55, .86]
DIS15	People make better consumer decisions if they don't believe advertising.	{1, Completely disagree}... {5, Completely agree}	3.64	1.17	.72	[.51, .85]

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
DIS16	**When I have to ask others for help, I...**	**{1, Never think about whether they might let me down}... {5, Always think about whether they might let me down}**	**3.02**	**1.17**	**.70**	**[.49, .83]**
DIS17	**I assume that promises will...**	**{1, Never be broken}... {5, Always be broken}**	**3.10**	**1.09**	**.66**	**[.43, .81]**
DIS18	Most people are...	{1, Not at all dishonest}... {5, Extremely dishonest}	2.46	1.01	.65	[.41, .80]
DIS19	The Internet as a whole is...	{1, Not at all unsafe}… {5, Extremely unsafe}	2.64	0.96	.61	[.36, .78]
DIS20	For the most part, people...	{1, Never take advantage of those who trust them}... {5, Always take advantage of those who trust them}	2.76	0.87	.59	[.33, .77]
DIS21	Most people are...	{1, Never exploited by those whom they trust}... {5, Always exploited by those whom they trust}	2.86	0.83	.51	[.22, .72]
DIS22	If they see the chance to take advantage of someone, most people would be...	{1, Not at all likely to try}... {5, Extremely likely to try}	3.00	1.16	.47	[.17, .69]
DIS23	In general, people are...	{1, Not at all unpredictable}... {5, Extremely unpredictable}	2.80	1.01	.47	[.17, .69]

Table 2. Items developed and piloted for the Disposition subscale during Study 1. Bolded items were retained through the beginning of Study 2.

Institutional features. In total, 25 out of 52 participants completed both administrations of the Institutional features subscale. The Pearson correlations between the first and second administrations for individual items ranged from .36 (95% CI [-.04, .66]) to .78 (95% CI [.56, .90]), averaging .56 (*sd* = .13). Three of these items were retained for the full scale. INS7 was removed due to very low retest reliability (*r* = .36); INS4 as it is conceptually similar to INS2 but performed worse in

terms of skew and retest reliability; and INS5 for low variability and strong positive skew. I retained INS1, INS2, INS3, and INS6 as the institutional features subscale. This four-item subscale features moderate internal consistency and performed reliably across the two-week delay between administrations: Cronbach's alpha was .72 (.74 standardized), and Pearson's *r* between the two administrations was .71 (95% CI [.44, .86]).

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
INS1	**I believe that Company] collects...**	**{1, Only the user data it needs}... {5, Considerably more user data than it needs}**	**3.00**	**1.36**	**.78**	**[.56, .90]**
INS2	**The stories I have heard about [Company] as a company are...**	**{1, Never negative}... {5, Always negative}**	**2.85**	**0.98**	**.64**	**[.33, .83]**
INS3	**Altogether, the policies and laws that protect people from harm by [Company] are...**	**{1, Not at all lacking}... {5, Extremely lacking}**	**2.29**	**1.33**	**.57**	**[.23, .79]**
INS4	When people talk about [Company], they...	{1, Never say bad things}... {5, Always say bad things}	2.82	1.07	.57	[.23, .79]
INS5	[Company] has...	{1, Never mistreated me}... {5, Always mistreated me}	2.33	1.06	.51	[.14, .75]
INS6	**[Company]'s policies for fixing issues that customers or users encounter are...**	**{1, Not at all inadequate}... {5, Extremely inadequate}**	**2.50**	**1.50**	**.47**	**[.09, .73]**
INS7	Choosing to rely on [Company] is...	{1, Not at all risky}... {5, Extremely risky}	2.47	1.22	.36	[-.04, .66]

Table 3. Items developed and piloted for the Institutional Features subscale during Study 1. Bolded items were retained through the beginning of Study 2.

Beliefs. In total, 34 out of 54 participants completed both administrations of the beliefs subscale. The Pearson correlations between the first and second administrations for individual items ranged from .34 (95% CI [.00, .61]) to .90 (95% CI [.81, .95]), averaging .77 (*sd* = .10). With a few notable exceptions, these items were highly reliable. I began with the most items in this subscale, as they constitute the core of distrust: the evaluation of an entity's incompetence, malevolence, or dishonesty. For the full scale, I removed or edited several items for having bimodal distributions, repetition, and to retain only unipolar items with a five-option response scale. I also removed 3 items (BEL16, BEL25, BEL35) that, upon reflection, were conceptually closer to measuring trust than distrust. The items I retained as the Beliefs subscale in the full scale were BEL9, BEL11, BEL18, BEL19, BEL27, BEL28, BEL32, and BEL36. This subscale has high internal consistency, with a Cronbach's alpha of .95, and the Pearson's correlation coefficient for this subscale between the first and second administrations was r = .90, 95% CI [.80, .95], showing high retest reliability for the subscale over two weeks.

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
BEL1	As a company, the quality of the work [Company] does is...	{1, Extremely good}... {7, Extremely bad}	3.26	1.97	.90	[.81, .95]
BEL2	[Company] makes the world...	{1, A much better place}... {5, A much worse place}	2.67	1.27	.88	[.77, .94]
BEL3	Relying on [Company] means that things will go...	{1, Extremely well}... {7, Extremely poorly}	3.52	2.05	.88	[.77, .94]

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
BEL4	I believe [Company] does...	{1, Much more good than harm}... {5, Much more harm than good}	2.50	1.28	.86	[.74, .93]
BEL5	If I could change how [Company] treats people, I would change...	{1, Nothing}... {5, Everything}	2.69	1.30	.86	[.74, .93]
BEL6	In general, I believe that [Company] is...	{1, Not at all misleading about its behavior}... {5, Extremely misleading about its behavior}	2.57	1.40	.86	[.74, .93]
BEL7	In order to do what they do, [Company]...	{1, Never exploits people}... {5, Always exploits people}	2.74	1.20	.85	[.72, .92]
BEL8	In my opinion, [Company] is...	{1, Extremely good at what they do}... {7, Extremely bad at what they do}	3.35	2.21	.84	[.70, .92]
BEL9	**[Company] is...**	**{1, Not at all incompetent}... {5, Extremely incompetent}**	**2.40**	**1.36**	**.83**	**[.68, .91]**
BEL10	[Company] seems...	{1, Not at all heartless}... {5, Extremely heartless}	2.41	1.50	.83	[.68, .91]
BEL11	**[Company]...**	**{1, Never tries to take advantage of people}... {5, Always tries to take advantage of people}**	**2.83**	**1.30**	**.83**	**[.68, .91]**
BEL12	As a company, [Company]...	{1, Never does a bad job}... {5, Always does a bad job}	2.76	0.95	.83	[.68, .91]
BEL13	I believe that [Company] is...	{1, Not at all insincere}... {5, Extremely insincere}	2.80	1.46	.82	[.67, .91]
BEL14	The claims that [Company] makes are...	{1, Not at all unreliable}... {5, Extremely unreliable}	2.46	1.37	.82	[.67, .91]
BEL15	[Company] is...	{1, Not at all corrupt}... {5, Extremely corrupt}	2.11	1.25	.81	[.65, .90]
BEL16	If they make a mistake, how confident are you that [Company] can fix it?	{1, Extremely confident}... {5, Not at all confident}	2.93	1.41	.81	[.65, .90]
BEL17	[Company] is...	{1, Not at all selfish}... {5, Extremely selfish}	2.78	1.38	.79	[.62, .89]
BEL18	**I believe that [Company]...**	**{1, Never tries to trick people}... {5, Always tries to trick people}**	**2.69**	**1.27**	**.79**	**[.62, .89]**

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
BEL19	**The company image that [Company] presents to the public is...**	**{1, Not at all fake}... {5, Extremely fake}**	**2.52**	**1.49**	**.79**	**[.62, .89]**
BEL20	How is [Company] at following through with its commitments?	{1, Extremely good}... {7, Extremely bad}	3.67	2.09	.78	[.60, .88]
BEL21	How often does [Company] act differently from the values they claim to hold?	{1, Never}... {5, Always}	2.93	1.10	.78	[.60, .88]
BEL22	I would describe the way [Company] acts as being...	{1, Not at all creepy}... {5, Extremely creepy}	1.91	1.19	.78	[.60, .88]
BEL23	In general, [Company] wants to make my life...	{1, Much better}... {5, Much worse}	2.56	1.19	.78	[.60, .88]
BEL24	Relying on [Company] is...	{1, A not at all unpredictable experience}... {5, An extremely unpredictable experience}	2.56	1.37	.78	[.60, .88]
BEL25	To what extent does [Company] try to help others outside the company?	{1, To a very large extent}... {5, To a very small extent}	3.33	1.23	.77	[.58, .88]
BEL26	[Company] is...	{1, Never dishonest}... {5, Always dishonest}	2.78	1.09	.76	[.57, .87]
BEL27	**I believe that [Company]...**	**{1, Never acts immorally}... {5, Always acts immorally}**	**2.56**	**1.04**	**.76**	**[.57, .87]**
BEL28	**[Company]'s values seem...**	**{1, The same as my values}... {5, Extremely different from my values}**	**2.78**	**1.34**	**.74**	**[.54, .86]**
BEL29	[Company] seems to be...	{1, Not at all deceitful}... {5, Extremely deceitful}	2.46	1.34	.73	[.52, .86]
BEL30	I believe that [Company]...	{1, Never acts against my best interest}... {5, Always acts against my best interest}	2.94	1.16	.73	[.52, .86]
BEL31	[Company]...	{1, Completely acknowledges what I want}... {5, Completely ignores what I want}	2.93	1.18	.72	[.50, .85]

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
BEL32	**How skeptical are you about whether [Company] is capable of doing a good job?**	**{1, Not at all skeptical}... {5, Extremely skeptical}**	**2.44**	**1.42**	**.72**	**[.50, .85]**
BEL33	When it comes to their work, [Company] is...	{1, Not at all careless}... {5, Extremely careless}	2.28	1.20	.71	[.49, .84]
BEL34	Compared to their competitors, [Company] is...	{1, Much less creepy}... {5, Much more creepy}	2.76	1.23	.69	[.46, .83]
BEL35	I believe that [Company]...	{1, Always keeps its commitments}... {5, Never keeps its commitments}	2.65	1.12	.68	[.44, .83]
BEL36	**To what extent does [Company] fall short at what it tries to do?**	**{1, Not at all short}... {5, Extremely short}**	**2.43**	**1.31**	**.68**	**[.44, .83]**
BEL37	There is something unethical about how [Company] acts.	{1, Completely disagree}... {5, Completely agree}	2.92	1.43	.66	[.42, .82]
BEL38	Relying on [Company] means that things will...	{1, Never go wrong}... {5, Always go wrong}	2.78	0.92	.65	[.40 .81]
BEL39	I feel like I know...	{1, None of [Company]'s values}... {5, All of [Company]'s values}	2.81	1.12	.34	[.00, .61]

Table 4. Items developed and piloted for the Beliefs subscale during Study 1. Bolded items were retained through the beginning of Study 2.

Intentions. In total, 29 out of 51 participants completed both administrations of the Intentions subscale. The Pearson correlations between the first and second administrations for individual items ranged from .50 (95% CI [.16, .73]) to .91 (95% CI [.82, .96], averaging .70 (*sd* = .14). With a few exceptions, these items were highly reliable. The reverse-coded items INT2 and INT6 were removed out of concern that they tapped into trust (intention to use) rather than distrust (intention to avoid). An

additional three items were removed for having substantially non-normal distributions, including severe skew or bimodal distributions. In the final subscale, I retained six items: INS1, INS5, INS7, INS9, INS10, and INS11. Internal consistency and retest reliability was high for this subscale: Cronbach's alpha of .90 (n=29), Pearson's *r* = .87 (95% CI [.74, .94]).

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
INT1	**In the future, I will...**	**{1, Certainly not avoid [Company]}... {5, Certainly avoid [Company]}**	**2.69**	**1.56**	**0.91**	**[.82, .96]**
INT2	How likely are you to rely on [Company] in the future?	{1, Extremely likely}... {5, Not at all likely}	3.27	1.55	0.89	[.78, .95]
INT3	If I needed something that several companies could provide at similar quality and cost...	{1, I would certainly go with [Company]}... {5, I would certainly not go with [Company]}	2.33	1.42	0.83	[.67, .92]
INT4	If you needed something that [Company] could do for you right now, but other companies could do with a short delay, would you use [Company] or wait for a competitor?	{1, Definitely use [Company]}... {5, Definitely avoid [Company]	3.02	1.39	0.83	[.67, .92]
INT5	**If avoiding [Company]'s products or services meant you would miss out on something you cared about, would you avoid or use their products or services?**	**{1, Definitely use}... {5, Definitely avoid}**	**2.78**	**1.17**	**0.73**	**[.50, .89]**
INT6	If a new product or service from [Company] received excellent reviews, how likely would you be to consider using it?	{1, I would certainly consider using it}... {5, I would certainly not consider using it}	2.9	1.15	0.66	[.39, .83]

Item number	Item stem	Response options	Mean	SD	Retest *r*	Retest *r* 95% CI
INT7	**If [Company]'s products or services were better than their competitors, how likely is it that you would go with a competitor over [Company]?**	**{1, I would certainly not go with a competitor}... {5, I would certainly go with a competitor}**	**2.65**	**1.23**	**0.65**	**[.37, .82]**
INT8	How likely are you to recommend that other people avoid using [Company]?	{1, Not at all likely}... {5, Extremely likely}	2.16	1.412	0.57	[.25, .77]
INT9	**If you wanted something that only [Company] could provide, would you use [Company] or do without?**	**{1, Definitely use [Company]}... {5, Definitely do without}**	**2.59**	**1.40**	**0.57**	**[.25, .77]**
INT10	**If competitors' products or services were more convenient than [Company]'s, how likely would you be to use a competitor instead of [Company]?**	**{1, I would certainly not use a competitor}... {5, I would certainly use a competitor}**	**3.63**	**1.11**	**0.56**	**[.24, .77]**
INT11	**How strongly do you feel about avoiding [Company] in the future?**	**{1, Not at all strongly}... {5, Extremely strongly}**	**2.41**	**1.40**	**0.50**	**[.16, .73]**

Table 5. Items developed and piloted for the Intentions subscale during Study 1. Bolded items were retained through the beginning of Study 2.

Discussion

In Study 1, I defined and created items for four survey subscales, each assessing a different aspect of distrust based on prior theoretical and empirical work. I conducted cognitive pretesting on these novel items, which improved the content validity of the subscales based on expert and non-expert feedback. Following the pretesting phase, I assessed the reliability for the proposed subscales. After removing extraneous and low performing items, each of the four subscales featured high retest

reliability (all *r*s > .70) across two weeks, comparing favorably to similar scales (Obermiller & Spangenberg, 1998; Weitzl, 2016). Three of the four subscales also feature high internal consistency, supporting that the items of each individual subscale converge on a single construct. The Institutional Features subscale had somewhat lower but still acceptable internal consistency (Cronbach's alpha = .72), which is unsurprising given that it is the shortest and most conceptually diverse construct of the four. There is one potential issue with internal consistency worth noting. The Beliefs subscale featured a Cronbach's alpha of .95, above the .90 threshold Streiner (2003) recommends as indicating potential issues with redundancy. Although internal consistency is frequently this high for trust and distrust subscales in other research [cite], there may be a methodological reason for it in the current study. The company elicitation prompts used in this study invited participants to rate the company they felt most positively or negatively about, increasing the likelihood that their attitudes This is a concern that I return to in the next study. With these key reliability features established, I move on to Study 2, a larger administration of the full scale to assess its factor structure and validity.

Study 2: Scale Validation

In Study 1, I generated and selected reliable, conceptually relevant distrust items for four subscales: Disposition to Distrust, Institutional Features of Distrust, Distrustful Beliefs, and Distrustful Intentions. I next administered the full scale to 400 US participants on Mechanical Turk to explore its factor structure and compare it to

the definitional model developed by McKnight and Chervany (2001). Additionally, I included prior scales assessing dispositional mistrust, brand trust, and brand satisfaction to establish convergent and discriminant validity for the newly-developed subscales.

Method

Participants. For the full administration of the scale, I recruited 400 participants using Mechanical Turk. MTurk workers who participated in Study 1 were excluded from participation so that this was the first time participants viewed these items. One participant completed the survey from outside the US, so their data was not included in the final analyses, n = 399. Participants were 41% female, 59% male, age ranging from 19 to 69 years old, average of 33 years old (*sd* = 9.5 years). This sample overrepresented ethnic majority participants: 80% of participants identified as White or European-American, 3% American Indian or Alaska Native, 8% Asian or Pacific Islander, 7% Black or African-American, 9% Hispanic Latino/Latina or Spanish, 7% selected multiple ethnicities. Educational attainment ranged from some high school to doctorate. 15% of the sample had a high school degree with no college, 26% some college but no degree, 14% associate's degree, 36% bachelor's degree, 9% had some level of graduate education, moderately oversampling those with some college or higher, while undersampling those with a high school degree compared to the general US population ("Population estimates, July 1, 2015, (V2015)," n.d.).

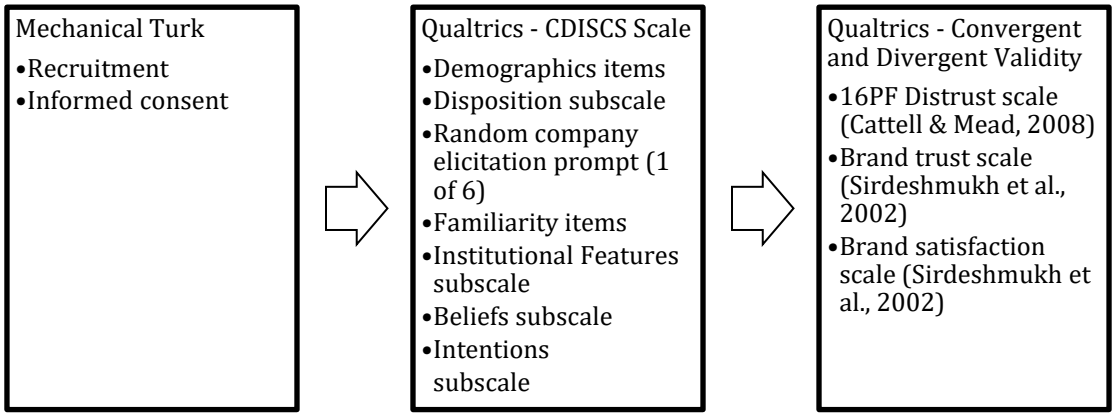

Figure 5. Flowchart showing survey administration procedure for Study 2. All CDISCS components were developed for the current study. Convergent and divergent validity scales were drawn from the cited works.

Procedure. Participants were recruited and provided informed consent as in Study 1. Due to the longer completion time for the full scale compared to the individual subscales from Study 1, participants in Study 2 were recruited to complete a 24-minute survey on their consumer beliefs and attitudes for $4. Most participants finished the survey in less time, M = 15 minutes, sd = 9 minutes.

After recruitment, participants answered demographic questions. Following demographics, participants completed the Disposition items in randomized order. Next, one company elicitation prompt was randomly presented from a pool of six company elicitation prompts. The six prompts included one from Study 1 that asked about a company the participant has stopped using, but asking participants to name the companies they have had the most positive or most negative experiences with may have polarized the responses. In Study 2, I deployed an expanded pool of prompts that contained less extreme wording:

> Thinking about companies whose online products, services, or website you have used in the past, what is the name of a company [(1) you have only dealt with once/(2) you have **stopped** using/(3)who you believe makes the world a better place/(4) who you believe makes the world a worse place/(5) you believe is good at what they do/(6) you believe is bad at what they do]? Please write only the name of the company in the space below. Take your time in answering this question, as the remainder of the survey will be about the company you name here.

After the company elicitation prompt, the name of the participant's chosen company was inserted automatically into the remaining questions on the survey. The prompts were generally effective in eliciting a variety of companies. Although the most frequent company chosen was Amazon.com, the online retailer who owns and operates Mechanical Turk (n = 61), 196 unique companies were elicited. These included large online companies (e.g., eBay, Facebook, Microsoft), retailers (e.g., Lush Cosmetics, Newegg, Walmart), companies providing various consumer services (e.g., AT&T, Chase Bank, Comcast), travel companies (e.g., Alaska Airlines, Iceland Air, JetBlue), and various smaller companies (e.g., Alloy Apparel, Cryptic Studios, Traffic Monsoon).

After the company elicitation prompt, all participants completed four items assessing their familiarity with the company. Most participants rated themselves as very-to-extremely familiar with the company they rated, and familiarity predicted less than 1% of subscale scores (all *r*s < .09). After the familiarity section, participants completed the Institutional Features, Beliefs, and Intentions subscales. Within each subscale, items were randomized to reduce order effects. Finally, three scales from prior literature were included to assess convergent and divergent validity for the

current scale: the Distrust subscale of the 16 Personality Factors Questionnaire (16PF) (Cattell & Mead, 2008), a clinically-focused measure of distrust that has been validated and widely used; and the Consumer Trust and Satisfaction scales developed by Sirdeshmukh, Singh, and Sabol (2002). In an independent evaluation of three brand trust measures commonly deployed in marketing research, Sirdeshmukh et al.'s was the only questionnaire that evinced unidimensionality as well as high reliability and validity (Brudvig, 2015).

I hypothesized that the 16PF Distrust subscale should correlate highly positively with the Disposition subscale I designed but not correlate strongly with the remaining three subscales. Conversely, I hypothesized that the Consumer Trust and Satisfaction scales would correlate highly negatively with the non-Disposition subscales but have no strong correlation with the Disposition subscale.

Analysis. The current scale was developed based on McKnight and Chervany's (2001) theoretical definition of distrust as comprised of four constructs that influence behavior: (1) the individual's disposition towards distrust, (2) institutional features that provoke distrust, (3) distrustful beliefs about the entity's incompetence, malevolence, or dishonesty, and (4) intentions to avoid the entity in the future. Study 1 showed that when considered independently, each of these four constructs appears to be reliably measurable with the items I developed. In Study 2, I aimed to validate the scale both in terms of how closely its features support the definitional model laid out by McKnight and Chervany, as well as on its own merits

as a reliable, valid, multidimensional instrument. To support those goals, I first conducted an EFA on the full scale to assess its factor structure. Although the disposition and intentions subscales were distinct, following McKnight and Chervany's hypothesized model, the institutional features and beliefs items interspersed, with two distinct, key elements of distrust emerging. The first emergent factor measures the respondent's assessment of the company's incompetence, whereas the second measures the respondent's assessment of the company's malevolence. After generating this new model, I analyzed the distributions of each subscale, and performed additional validity and reliability testing based on the full administration of the survey on Mechanical Turk.

Visually examining the scree plot (Figure 6) shows that there are two high eigenvalue factors, with the third and fourth factors having eigenvalues around 1. Because this scale was theoretically driven by a four construct model of distrust, I extracted four factors in the following analyses. However, I encountered high correlations between three of the factors that led us to generate two- and three-factor models for the data to see whether they would be more interpretable. The three-factor model extracted a single high-loading "distrust" factor containing items from the Institutional Features, Beliefs, and Intentions subscales, a distinct "Disposition to Distrust" factor containing only the Disposition subscale items, and a third factor that had low (<.4), scattered loadings from a small number of items. The two-factor modelI generated was similar in the distinction it made between the factors: all of the

non-Disposition items loaded highly (>.6) on one factor, and the Disposition items all loaded highly on the other factor (>.6). This model cleanly separated the disposition subscale from the remaining items, but the composite distrust factor would not be useful in assessing the relationships between components of distrustI hypothesized from McKnight's definitional constructs. As a result, I continue with the four-factor model for the following analyses and results.

I hypothesized that there would be differences in the distributions of average scores for each subscale. Because the Disposition subscale was designed to capture a personality trait, I hypothesized that its scores would be approximately normally distributed. By contrast, I expected substantial positive skew for the other three subscales as a result of the company elicitation prompts, which prompted a large proportion of the participants to measure their distrust for a company they felt strongly positively about. I did, however, expect greater skew for the Incompetence and Intentions subscales compared to the Malevolence subscale. Consumers generally expect companies to act selfishly to some degree, and I expected that to manifest as a higher mean score and less positive skew for the Malevolence subscale than the Incompetence and Intentions subscales.

Additionally, within the four-factor structure, I tested specific hypotheses about the relationships between subscales drawn from McKnight and Chervany (2001). The Disposition subscale should be moderately positively correlated with the other three subscales, as the predisposition towards distrust should affect

interpretations of interactions with companies and the general tendency to avoid relying on a company. The Incompetence and Malevolence subscales should be highly positively correlated with the Intentions subscale, because those attributions should form the basis of decisions the respondent makes to avoid the company. Finally, Incompetence and Malevolence should be distinguishable from each other, with $r < .70$ pairwise correlations between them. Although prior trust and distrust research has often found high correlations between constructs related to these subscales (e.g., Cho, 2006), the current survey was designed to be better able to effectively discriminate between these constructs.

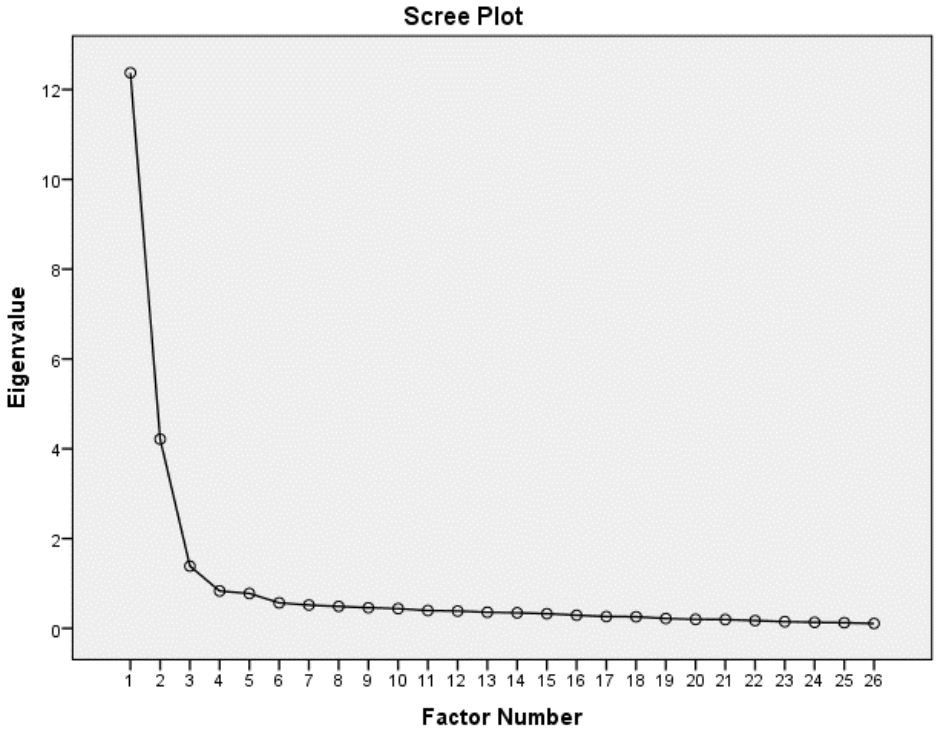

Figure 6. Scree plot showing Eigenvalues of the factors extracted from Exploratory Factor Analysis of the full CDISCS scale in Study 2.

Results

Overall, the scale appears to have high reliability and validity with a few important caveats. Although confirmatory factor analyses show that the four-factor model described is a good fit for the data, there is substantial overlap between items in the incompetence and malevolence subscales, evidenced by moderately high correlations between items in these subscales, and between the subscales themselves.

Factor analysis. I conducted Exploratory Factor Analysis (EFA) on the entirety of the proposed scale using direct oblimin rotation. Oblique rotations such as direct oblimin are recommended over orthogonal rotations for EFA when there are high hypothesized correlations between the factors, as the pattern matrix output from oblique rotations show the unique contributions to each factor absent the common contributions that the structure matrix from an orthogonal rotation would include. Initial EFA showed that only two items had communalities <.5, indicating that their low factor loadings justify removal: DIS11 (.464), and INS11 (.292). Returning to the Cronbach's alphas calculated in Study 1, removing DIS11 would only slightly decrease Cronbach's alpha from .93 to .92, and removing INS11 would increase Cronbach's alpha for that subscale from .72 to .74. As a result, both items were removed from the scale and the remaining analyses. The remaining items comprised the full scale deployed in Study 2 and Study 3 (full item stems and response options are included in Appendix 1).

Subscale	Variance Explained	Item	Factor Loading	Communality
Disposition	16.26%	DIS1	.753	.584
		DIS2	.711	.540
		DIS3	.721	.520
		DIS4	.763	.590
		DIS5	.756	.567
		DIS6	.765	.576
Incompetence	5.46%	INC1	.873	.837
		INC2	.759	.868
		INC3	.784	.809
		INC4	.540	.579
Malevolence	50.26%	MAL1	.749	.753
		MAL2	.971	.852
		MAL3	.932	.822
		MAL4	.922	.833
		MAL5	.726	.780
		MAL6	.584	.567
		MAL7	.561	.633
Intentions	3.38%	INT1	.781	.848
		INT2	.645	.821
		INT3	.926	.698
		INT4	.663	.641
		INT5	.731	.733
		INT6	.850	.705

Table 6. Details of Exploratory Factor Analysis from full CDISCS scale deployment in Study 2. Item numbering corresponds to the items included in Appendix 1.

As previously mentioned, this empirical analysis grouped the subconstructs of distrust somewhat differently from the definition McKnight and Chervany (2001) proposed. The Disposition and Intentions subscales do appear to echo the constructs of the same name from McKnight and Chervany's definitional model. However, rather than the items assessing the company's capabilities and tendencies separating based on whether they are institutional features or consumer beliefs, they comprised two different subdimensions that I suggest reflect the respondent's assessment of the company's (1) incompetence and (2) malevolence and dishonesty. Incompetence here includes items on skepticism about the company's ability to come through, how far the company falls short, its policies for fixing issues consumers encounter, and a direct question about its incompetence. Malevolence includes the items that assess malevolence and dishonesty from the institutional features and beliefs items, such as the company's propensity to take advantage of others, its fakeness and immorality, as well as the negativity of its reputation. The total variance explained by the four factors was 75.36%, with the largest proportion of variance explained by this model coming from the Malevolence subscale (50.26%). However, there is reason to suspect this is not a particularly meaningful statistic. Prior surveys that include highly-correlated factors often appear to have factors switch between high (>40%) and low (<5%) variance explained when analyses appear otherwise similar (Weitzl, 2016). This reflects the high overlap between the combined factors rather than poor performance in one or the other factor.

Subscale	Average Inter-Item Correlation	Cronbach's Alpha (standardized)	Item	Item-to-Total Correlation	Mean (SD)
Disposition	.556	.88 (.88)	DIS1	.71	2.85 (1.07)
			DIS2	.68	2.70 (1.00)
			DIS3	.67	2.85 (1.06)
			DIS4	.70	3.09 (.84)
			DIS5	.70	3.01 (.85)
			DIS6	.69	2.85 (.83)
Incompetence	.764	.93 (.93)	INC1	.86	2.20 (1.28)
			INC2	.88	2.41 (1.40)
			INC3	.86	2.39 (1.33)
			INC4	.73	2.50 (1.41)
Malevolence	.737	.95 (.95)	MAL1	.84	2.79 (1.36)
			MAL2	.88	2.75 (1.33)
			MAL3	.88	2.56 (1.14)
			MAL4	.87	2.53 (1.28)
			MAL5	.85	2.46 (1.41)
			MAL6	.74	2.44 (1.31)
			MAL7	.78	2.79 (1.09)
Intentions	.729	.94 (.94)	INT1	.88	2.77 (1.58)
			INT2	.85	2.56 (1.56)
			INT3	.79	2.43 (1.38)
			INT4	.78	2.67 (1.44)
			INT5	.83	2.83 (1.43)
			INT6	.81	2.72 (1.39)

Table 7. Inter-item correlation, internal consistency, corrected item-to-total correlations, and mean item score results from Study 2 CDISCS scale deployment.

Next, I analyzed the reliability of each subscale, including internal consistency and item characteristics. The factors that emerged from the EFA show higher internal consistency than those shown in Study 1, exposing a potential validity concern with the current version of the scale. Cronbach's alpha, a coefficient assessing the proportion of variance captured by the items if used as a scale, was over .88 for each subscale, indicating each one is measuring a coherent construct. Cronbach's alpha of .70 or higher is considered desirable to ensure that the items hang together appropriately within each subscale (Tavakol & Dennick, 2011), however the Incompetence, Malevolence, and Intentions subscales' alphas were over .90 and approach .95, past the threshold considered to be an indicator of a scale's redundancy or over-unidimensionality (Streiner, 2003). The average interitem correlations for these three subscales are also high, (Incompetence: .76, Malevolence: .74, Intentions: .73). One limitation of the current administration is that the company elicitation prompts may have artificially depressed variability in responses. Because participants were asked to name a company rather than being assigned a company to assess, the companies that were elicited may have been the most salient due to strong positive or negative associations with the company. Participants may have been induced to pick companies that they felt strongly positively or strongly negatively about, inflating the interitem correlations and Cronbach's alpha for the subscales. I return to this in the general discussion.

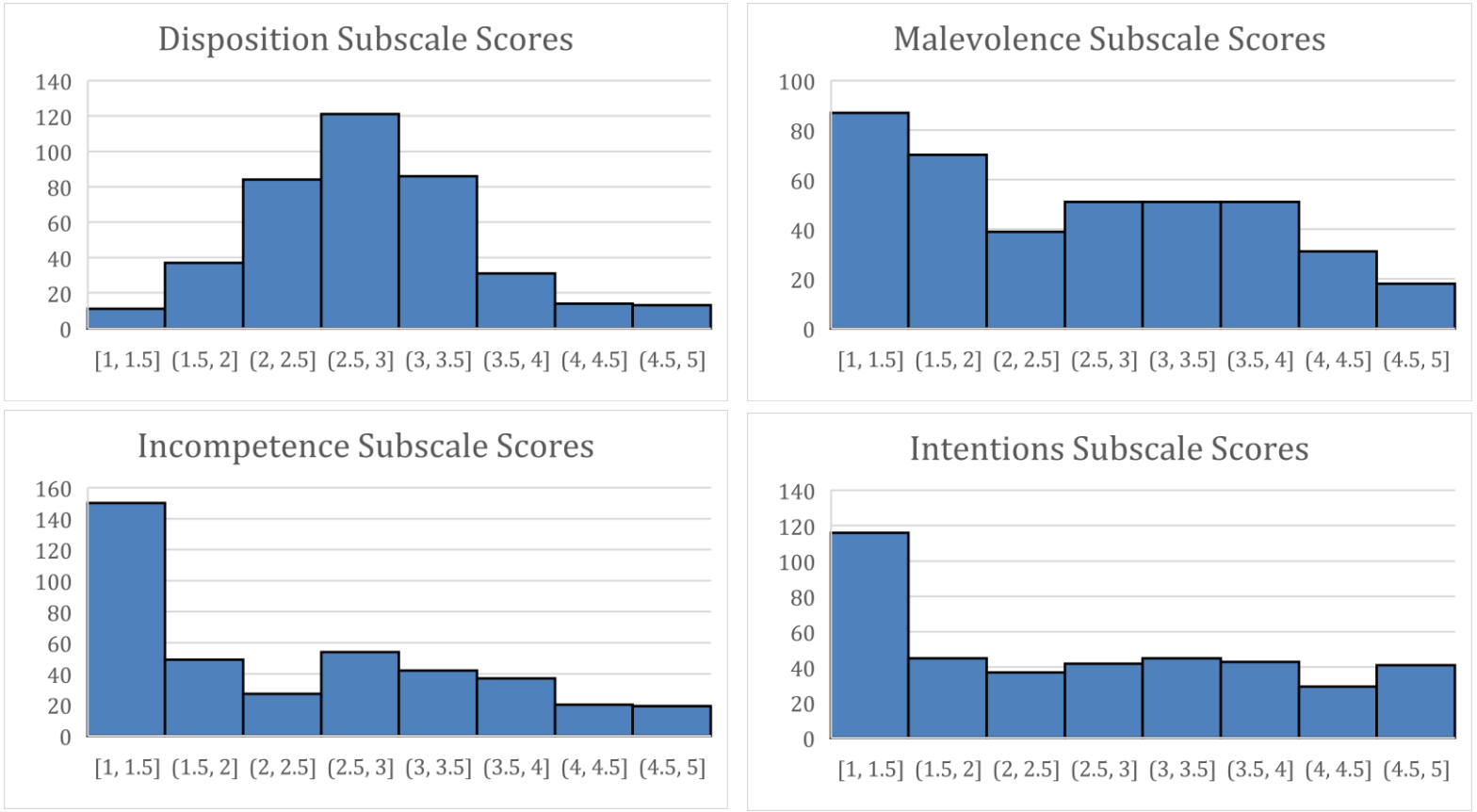

Figure 7. Frequency histograms showing distributions of subscale scores for each of the four CDISCS subscales.

Descriptive statistics. Within each subscale, individual item scores were averaged to obtain estimates of their central tendency and variability. Each subscale score therefore ranged from 1 to 5, with a score of 1 representing the lowest possible distrust and a score of 5 representing the highest possible distrust: the individual's disposition towards distrust for the Disposition subscale, and their distrust for the company for the other three subscales. I remind the reader that because of the company elicitations prompts used, only the Disposition subscale was hypothesized to have a normal distribution.

Subscale	Mean (SD)	95% CI Mean
Disposition	2.89 (0.75)	[2.82, 2.97]
Incompetence	2.38 (1.23)	[2.26, 2.50]
Malevolence	2.62 (1.12)	[2.51, 2.73]
Intentions	2.66 (1.29)	[2.54, 2.80]

Table 8. Means, standard deviations, and 95% confidence intervals for the mean of each subscale from Study 2.

As shown in Table 8, each subscales' average score was below 3, the midpoint of the scale. The distributions for each subscale matched my hypotheses, as seen in Figure 7. Disposition appears to have an approximately normal distribution, with Malevolence showing somewhat less positive skew (skewness = .28) compared to the Incompetence (skewness = .44) and Intentions (skewness = .28) subscales. Disposition to distrust is a personal trait, and therefore the normal distribution of responses supports the subscale's ability to capture normal variability in this trait. Two of the six company elicitation prompts asked participants to rate a company who

they feel does great work or makes the world a better place, and those participants largely rated the target company's Incompetence, Malevolence, their own Intentions to avoid the company below 2/5, indicating very low distrust. Importantly, despite this skew, the full 1-5 range of scores was represented for all subscales. This supports the ability of each subscale to measure a wide spectrum of distrustful beliefs and attitudes.

Subscale	Disposition	Incompetence	Malevolence	Intentions
Disposition	1	.16 [.06, .25]	.11 [.01, .21]	.07[-.03, .17]
Incompetence		1	.81 [.77, .84]	.79 [.75, .82]
Malevolence			1	.76 [.72, .80]
Intentions				1

Table 9. Inter-scale correlations between CDISCS subscales, 95% confidence intervals for correlation coefficients in parentheses.

Inter-scale correlations. My theory-driven hypotheses about the relationships between subscales received mixed support in this study. In accordance with McKnight and Chervany's (2001) definitional model, the Intentions subscale was highly correlated to the Incompetence and Malevolence subscales. However, Disposition's positive correlations with Incompetence (r = .16) and Malevolence (r = .11) were small. This contradicts McKnight and Chervany's model, which predicted a close relationship between a person's disposition towards distrust and their evaluations of distrust for other entities. The use of company elicitation prompts may have depressed the effect of dispositional distrust on participants' distrust assessments of the companies, but it is also likely that prior experience plays a larger role than

disposition on distrust for a company. Finally, I mention that the inter-scale correlations between Incompetence, Malevolence, and Intentions were higher than anticipated, all *r*s > .76. Although not unusual for trust and distrust surveys (e.g., Weitzl, 2016), these high correlations suggest there is significant conceptual overlap between the latent variables each subscale purports to measure. Some overlap is expected as these three subscales all assess aspects of distrust for the target company, but the degree of overlap is a potential issue that I return to in the general discussion.

Confirmatory factor analysis. Because the Incompetence, Malevolence, and Intentions subscales appear to be highly intercorrelated, I conducted Confirmatory Factor Analysis (CFA) to test whether the data suggest distrust is a second-order latent variable that underlies the first-order factors identified in the EFA. By comparing a model featuring this second-order latent variable to one without it, I hoped to establish that the four subscales of the CDISCS do, in fact, tap into a larger Distrust construct. This sample of nearly 400 participants exceeds the minimum sample size recommended for CFA in prior literature, as the sample has over 10 observations per parameter and more than 200 participants total (Hair, Black, Babin, Anderson, & Tatham, 2010; Jöreskog & Sörbom, 1989). The items from the four subscales were entered into two maximum-likelihood CFA models that were compared: in the first model (M1) seen in Figure 8, a latent variable was included for each subscale to reflect its items; in the second model (M2) seen in Figure 9, distrust

was added as a second-order latent variable that connected to each subscale's latent variable.

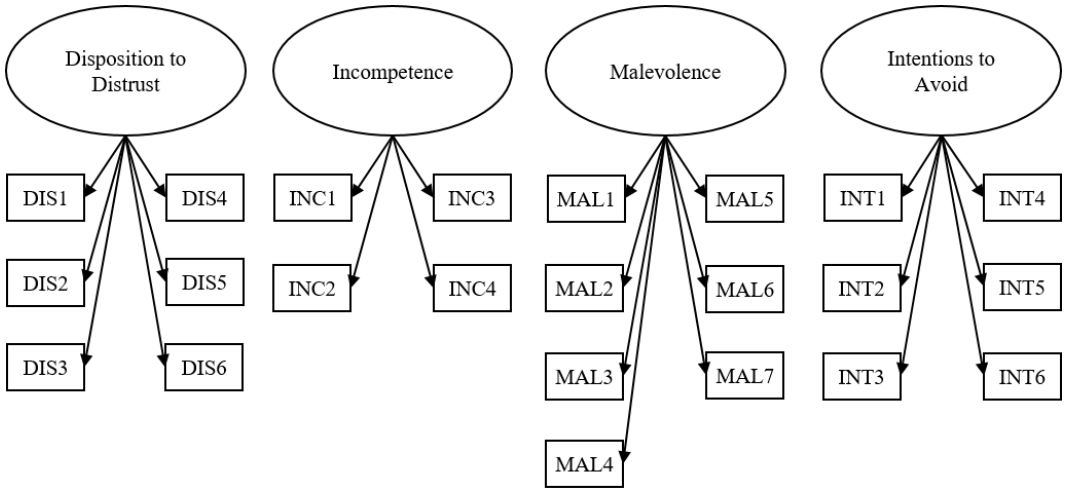

Figure 8. Diagram showing the Confirmatory Factor Analysis model for Model 1, in which four first-order factors were included. Independent error terms specified for each parameter not shown.

The models were compared using model fit, CFI, RMSEA, p_{close}, and AIC. Model fit for each model was assessed using a chi-square test. Both models fit the dataset (χ^2_{M1} (231)= 1335.03, $p < .001$; χ^2_{M2} (226)= 474.42, $p < .001$), with M2 outperforming M1 as evidenced by the lower chi-square value. That held for the other fit indices as well. CFI, the Comparative Fit Index, ranges from 0 to 1, with values over .95 indicating excellent fit. Unlike other indicators of model fit, RMSEA, the Root Mean Square Error of Association, corrects for the number of parameters in the model to prevent more complex models from outperforming simpler models that are otherwise equivalent in fit. RMSEA between .05 and .1 suggests close model fit, with p_{close} of >.05 supporting that the true RMSEA value is below .05.

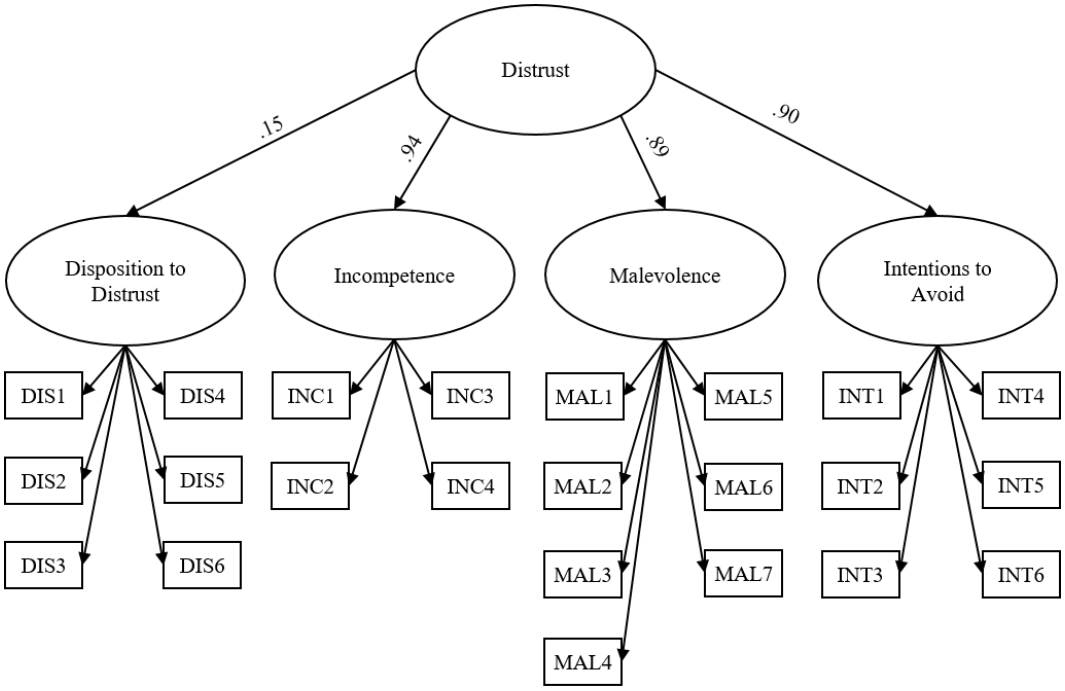

Figure 9. Diagram showing the Confirmatory Factor Analysis model for Model 2, which includes a second-order latent variable representing overall distrust for the specific company. Estimated correlations between the first- and second-order latent variables shown. Independent error terms specified for each parameter not shown.

AIC, the Akaike information criterion, allows for comparisons of different models' fit to the same dataset proportional to the number of parameters estimated, similar to the RMSEA. When AIC differs between models, the model with the lower AIC is preferred (Tabachnick & Fidell, 2007). The model containing the second-order distrust latent variable (CFI = .97, RMSEA = .053, p_{close} = .251, AIC = 620.42) outperformed the model without the second-order distrust variable (CFI = .87, RMSEA = .110, p_{close} < .001, AIC = 1471.03). This provides evidence of construct validity for the CDISCS scale. Treating each subscale as a subconstruct of distrust produces a better fit to the collected data than treating each subscale as a separate

construct. Having described the general features of the subscales and their related factor models, I now turn to an important aspect of their reliability: differential reliability, the scale's performance across different demographic groups.

Demographics analysis. I analyzed for differences in subscale performance based on demographic information collected from respondents, including gender, age, ethnicity, and education. In the following analyses, differences in group sizes led to significantly different variances, violating an assumption underlying the standard Student's *t* test. For consistency, I therefore use Welch's *t*-test for the following comparisons, which accommodates unequal variances between groups. Subscale average scores for men and women were compared by unequal variances t-tests, showing that the Disposition, Incompetence, and Intentions subscales performed

Subscale	Male n	Male mean (SD)	Female n	Female mean (SD)	*t* (unequal variances)	*p*
Disposition	233	2.93 (0.78)	164	2.84 (0.70)	1.22 (df = 372)	.233
Incompetence	234	2.45 (1.25)	164	2.27 (1.19)	1.45 (df = 361)	.149
Malevolence	233	2.73 (1.08)	165	2.46 (1.17)	2.23 (df = 334)	.024
Intentions	233	2.75 (1.23)	165	2.54 (1.36)	1.60 (df = 332)	.112

Table 10. Results from unequal-variances t-test comparing male and female subscale average scores for each subscale of the CDISCS.

similarly for men and women, (all *t*s < 1.60, all *p*s > .112) Male participants rated their target companies as a quarter-point more malevolent than female participants, which is a significant (p = .024) but negligible difference overall. .

To assess whether the subscales perform differently based on age, I calculated Pearson's correlation coefficients between participant age and each of the subscale scores. Incompetence and Intentions scores appeared not to be linearly related to age, suggesting those two subscales performed similarly for participants regardless of age. The Disposition subscale did show a small-to-moderate negative correlation with age (r = -.21, 95% CI [-.30, -.11]), which aligns with prior research showing that younger individuals tend to have a more distrusting mindset which decreases with age (Fett, Gromann, Giampietro, Shergill, & Krabbendam, 2014; Sutter & Kocher, 2007). Malevolence showed a trend towards decreasing with age, but the upper end of the 95% confidence interval for that correlation was sufficiently close to zero to suggest this effect is likely small.

Subscale	n	Correlation with age (*r*)	95% CI for *r*
Disposition	397	-.21	[-.30, -.11]
Incompetence	398	-.08	[-.18, .02]
Malevolence	398	-.10	[-.20, .00]
Intentions	398	-.06	[-.16, .04]

Table 11. Correlations between participant age and subscale average score for subscales of the CDISCS.

The Mechanical Turk sample undersampled ethnic minority participants, so I was unable to conduct inferential statistics analyzing each ethnic group separately.

Instead, I compare the confidence intervals for the mean of participants who identified as White/European-American and those who identified as a member of a non-White ethnic minority group. Because several participants identified as both White and part of an ethnic minority group, these groups have overlapping members. As seen in Table 12, Disposition to Distrust score was the only subscale of the four where means for either group lay outside the confidence interval of the other group. Although the 95% CI for the two groups did overlap slightly (White: [2.77, 2.93], Ethnic minority: [2.88, 3.18]), ethnic minority participants appear to have a somewhat higher disposition to distrust compared to the White/European-American participants. In fact, the ethnic minority disposition to distrust was the only subscale mean I observed that was higher than the scale midpoint of 3.

I had hypothesized that if any subscale differences based on ethnicity existed, they would appear in the Malevolence subscale, as ethnic minority consumers are more likely to encounter discrimination in consumer settings that may inform their interpretations of companies' motives (e.g., Warshaw, Taft, & Woodruff, 2016). The fact that no such differences showed up in this study suggests that either the instrument is not sufficiently sensitive to discriminatory aspects of company behavior, the sample had insufficient ethnic minority participants to find a difference in response patterns that does exist, or the companies elicited by the prompts were not companies who participants feel engage in that type of behavior. It is encouraging that the subscales assessing distrust in a company appear to be equivalently

measuring distrust across this coarse comparison of ethnic majority or minority participants, but additional administrations that recruit more ethnic minority participants are necessary to confirm that group differences were not masked by the current recruitment or grouping strategy.

Subscale	W/EA n	W/EA mean (SD)	W/EA 95% CI Mean	EM n	EM mean (SD)	EM 95% CI Mean
Disposition	315	2.85 (0.72)	[2.77, 2.93]	103	3.03 (0.77)	[2.88, 3.18]
Incompetence	315	2.34 (1.21)	[2.21, 2.48]	103	2.41 (1.23)	[2.17, 2.65]
Malevolence	315	2.62 (1.13)	[2.49, 2.74]	103	2.51 (1.10)	[2.30, 2.72]
Intentions	315	2.65 (1.28)	[2.50, 2.79]	103	2.61 (1.32)	[2.35, 2.87]

Table 12. Comparison of White/European-American (W/EA) and ethnic minority (EM) participants' average subscale scores, including 95% confidence intervals.

Finally, I looked for differences in subscale means based on education. As with ethnicity, Study 2's participants mostly came from one of two groups: those with a bachelor's degree or higher, and those with a high school degree but some or no college. I therefore tested whether there were significant differences in subscale means based on the presence or absence of a college degree, using unequal variances *t*-tests. Of the subscales, only Disposition showed a significant difference based on education: participants with no college degree showed a significantly higher disposition towards distrust compared to participants with a college degree or higher. I note, however, that the 95% CI of the difference between the means was [.01,.32] suggesting this difference is small, at most.

Subscale	College+ n	College+ mean (SD)	No College degree n	No College degree mean (SD)	*t* (unequal variances)	*p*
Disposition	232	2.82 (0.72)	165	2.99 (0.78)	2.15 (335)	.032
Incompetence	233	2.35 (1.21)	165	2.41 (1.25)	0.42 (346)	.673
Malevolence	232	2.66 (1.13)	166	2.56 (1.11)	0.94 (359)	.347
Intentions	233	2.67 (1.30)	165	2.66 (1.28)	0.07 (357)	.947

Table 13. Results from unequal-variances *t*-tests comparing subscale average scores of participants with a college degree or higher to participants with no college degree.

Convergent and divergent validity. Finally, I included three scales from previous literature in this administration to test the convergent and divergent validity of the four subscales: Cattell's 16 Personality Factors Distrust (16PF) subscale (Cattell & Mead, 2008), and the Brand Trust and Satisfaction subscales from (Sirdeshmukh et al., 2002). Pearson's correlation coefficients were calculated between for each of the four CDISCS subscales to the three external subscales (see Table 14). Each of my hypotheses were supported: the Disposition subscale correlated highly positively with the 16PF Distrust subscale, and slightly negatively with the Brand Trust and Satisfaction subscales. The Incompetence, Malevolence, and Intentions subscales correlated highly negatively with the Brand Trust and Satisfaction subscales, and slightly positively with the 16PF Distrust subscale. These findings support the validity of the subscales by establishing convergent and divergent validity with prior scales assessing similar and dissimilar constructs.

Subscale	16PF n	16PF *r*	Brand Trust n	Brand Trust *r*	Brand Satis. n	Brand Satis. *r*
Disposition	370	.70 [.64, .75]	385	-.11 [-.21, -.01]	387	-.12 [-.22, -.02]
Incompetence	371	.22 [.12, .31]	386	-.89 [-.91, -.87]	388	-.83 [-.86, -.80]
Malevolence	371	.16 [.06, .26]	387	-.84 [-.87, -.81]	389	-.75 [-.79, -.70]
Intentions	372	.14 [.04, .24]	386	-.85 [-.88, -.82]	388	-.84 [-.87, -.81]

Table 14. Correlations between CDISCS subscales and three scales from prior literature included in the Study 2 administration to assess convergent and divergent validity of CDISCS. 95% confidence intervals for correlations between subscales included in parentheses.

Study 3: Criterion Validity Study

During analysis of Study 2, an opportunity arose to evaluate the criterion validity of the CDISCS by incorporating news about a newly-uncovered scandal into a new deployment of the scale. Study 3 describes that deployment, an intervention study which was structured to assess the CDISCS' ability to assess change in a respondent's distrust for a company due to a naturalistic, distrust-enhancing intervention. In September 2016, a scandal arose that showed employees of the US bank Wells Fargo had been fraudulently opening bank and credit accounts for customers without their consent between 2011 and 2015, to artificially inflate employee performance metrics and revenue by assessing account fees. The bank was fined $185 million by the US Consumer Financial Protection Bureau, and 5,300 employees were fired for participating in this fraud. This provided an opportunity for a natural test of the CDISCS' ability to measure change in a person's distrust for a company. My expectation was that learning about this scandal would be highly likely

to increase distrust, which I tested using a within-subjects intervention. I therefore recruited Wells Fargo customers who completed the scale both before and after learning about the scandal, and I assessed changes in their distrust using the scale carried forward from Study 2.

Method

News broke regarding the Wells Fargo scandal on September 8, 2016 when they were fined by the US Consumer Financial Protection Bureau. On September I administered a short screener to 380 Mechanical Turk workers ($.03 for < 1 minute survey) in the US to screen for prospective participants who both (1) were current or previous Wells Fargo accountholders, and (2) had not heard about the Wells Fargo scandal. I filtered these participants using two items on the screener: "Which of the following banks do you currently have at least one account with? [J.P.Morgan Chase, Bank of America, Wells Fargo, Citigroup,…Other, I do not have a bank account]" and "Have you heard any news about your bank in the last seven days? [Yes, No]." In order to prevent respondents from being able to guess the purpose of the screener, the screener also included five distractor items that did not relate to banking, e.g., "What is your current cell phone's operating system?", "Have you heard any news about a new cell phone operating system release in the past seven days? [Yes,No]". Of the 380 screened participants, 70 reported being current or past Wells Fargo accountholders, and 27 of those claimed to have not heard any news about their banks in the past week. Those 27 participants were recruited as participants in the

intervention study using Mechanical Turk ($3 for an 18-minute study), with 25 participants participating in the procedure described next.

The intervention study procedure included three phases: pre-test, intervention, and post-test. The pre-test included the demographic questions and four subscales as reported in Study 2 with Wells Fargo as the target company for all participants, preceded by a few additional questions asking for the participant's familiarity and knowledge about Wells Fargo. In the intervention, each participant was directed to click a provided link and read the news article on that page in its entirety before continuing. The article was "Wells Fargo Fined $185 Million for Fraudulently Opening Accounts" (Corkery, 2016), a 1,126 word summary of the Wells Fargo scandal. An image of the article's text was made available for participant who were unable to view the article at the provided link, but all 25 participants reported being able to access and read the article using the link. To ensure the validity of these findings, after reading the article, participants were asked whether they had heard the story contained in the article prior to the study, and they were assured that their compensation would not be affected based on their answer to the question. Two participants reported that they had previously heard the story, so those participants did not complete the post-test phase. The remaining 21 participants were asked to rate how negative or positive the story was about Wells Fargo, and how credible they viewed the New York Times as a source of news, and then they completed the survey again as the post-test measure except for the Disposition scale, which was not

included to reduce survey length and strain on the participants. One additional participant did not complete the second half of the survey, so their data were not included in the following analyses. Finally, as a manipulation check, I asked the participants a single open-ended question: "How have your feelings about Wells Fargo changed as a result of learning about this news story?" The responses to this question confirmed that the overwhelming majority of participants were surprised at Wells Fargo's actions and felt greater mistrust for the bank after reading the article, e.g., "It has made me more cautious of them", although several participants did note a general distrust for banks that reduced the impact of this news, e.g., "I know banks sometimes do things they shouldn't".

I treated both administrations of the survey as opportunities to test hypotheses about the subscale scores as assessments of known-groups validity. The pre-test provided an estimate of Wells Fargo customers' existing distrust for the company. I hypothesized that as current customers of the bank, participants' pre-test Incompetence, Malevolence, and Intentions scores would be substantially lower than the average scores I observed for these subscales in Study 2. I also hypothesized that the Incompetence, Malevolence, and Intentions scores would increase moderately between the pre-test and post-test. I expected the highest difference to be for Malevolence as the news story was about the company taking advantage of consumers. However, the Incompetence items assess skepticism that the company is able to accomplish its goals effectively, so I did expect some increase in this

subscale's scores. I expected the lowest difference for the Intentions scale, as although the story may provoke negative sentiment toward Wells Fargo, changing banks may appear a drastic step to customers who believe the issue has already been addressed within the company after being made public.

Results

The data collected in this intervention study supported the hypotheses regarding differences from the sample drawn in Study 2 and pre-test to post-test differences based on the intervention. The participants recruited for this study rated themselves, on average, as moderately-to-very familiar with Wells Fargo as a company (m = 3.91/5, sd = 0.90) and slightly-to-moderately familiar with Wells Fargo's policies (m = 2.78/5, sd = 1.17). Of the 21 participants, 19 had been customers of Wells Fargo for at least two years, and 13 of those for over five years. I calculated pre-test and post-test scores for each of the three subscales for each participant by averaging the 1-5 responses for the items in each subscale. I then compared the mean scores for the Incompetence, Malevolence, and Intentions subscales of the participants in this study to those in Study 2. Each subscale's mean pre-test score in Study 3 was at least 0.3 points below its Study 2 equivalent, as seen

Subscale	Study 2 Mean	Pretest Mean (SD)	Posttest Mean (SD)	Mean Difference (SD)	Mean Difference 95% CI	$t(20)$	p
Incompetence	2.38	1.67 (0.90)	2.78 (1.10)	1.11 (0.67)	[0.80, 1.41]	7.59	<.001
Malevolence	2.62	2.31 (0.98)	3.05 (0.85)	0.75 (0.46)	[0.54, 0.96]	7.38	<.001
Intentions	2.67	1.87 (0.94)	2.70 (1.18)	0.82 (0.76)	[0.47, 1.16]	4.95	<.001

Table 15. Results from paired t-tests conducted as follow-up analyses to assess the change in subscale average scores before and after a distrust-enhancing intervention.

in Table 15. As expected, participants evidence lower distrust for a bank whom they have an account with compared to the participants in Study 2 rating a mix of companies.

A Repeated Measures MANOVA was conducted to test whether the intervention was effective in increasing distrust, which would support the scale's ability to measure such a difference. The pre-test and post-test scores for Incompetence, Malevolence, and Intentions were entered into this MANOVA. Distrust increased significantly from pre-test to post-test on each of the three subscales, and overall (Wilk's Lambda = .219, F (3,18) = 21.43, p < .001). The intervention explained over three-quarters of the variance explained by the change from pre-test to post-test (partial-eta squared = .78). On average, Incompetence subscale ratings increased on average by 1.11 points (sd = 0.67, t(20) = 7.586, p < .001; 95% CI [0.80,1.41]); Malevolence subscale ratings increased on average by 0.75 points (sd = 0.46, t(20) = 7.38, p < .001; 95% CI [0.54, 0.96]); and Intentions scores increased on average by 0.82 points, t(20) = 4.95, p < .001; 95% CI [0.47, 1.16]). The scale, therefore, appears to adequately measure changes in distrust that result from learning about a scandal involving a company that respondents have interacted with previously.

Subscale	Reliability			Validity				
	Retest (2 weeks between)	Differential across demographics	Internal Consistency	Face	Content	Convergent	Discriminant	Criterion
Disposition	High	Medium	High	High	High	High	High	Untested
Incompetence	High	High	Too high	High	Unclear	High	Low	High
Malevolence	High	Medium	Too high	High	Unclear	High	Low	High
Intentions	High	High	Too high	High	Unclear	High	Low	High

Table 16. Table summarizing the reliability and validity of each of the four subscales of the CDISCS scale following the three studies described in this thesis.

General Discussion

I present the Consumer Distrust in Specific Company Scale (CDISCS), a novel survey instrument that assesses an individual's distrust for a company, which flexibly accommodates a wide spectrum of companies. The development and validation of the CDISCS drew heavily from existing work that proposed new models of trust and distrust (e.g., Komiak & Benbasat, 2008; Lankton & McKnight, 2011; D. H. McKnight & Chervany, 2001), and this thesis provides a methodological tool to explore those models empirically in the future.

As shown in Table 16, the current version of the scale features high retest reliability (Study 1) and evidences high criterion validity evidenced by sensitivity to a naturalistic intervention (Study 3). The Disposition subscale appears to have excellent internal consistency, reliability, convergent validity, and construct validity. Although the current study did not provide criterion validity for this subscale, it appears to be otherwise reliable and valid. The Incompetence, Malevolence, and Intentions subscales require slightly more explanation to describe their current status. Each of these subscales showed high reliability, as well as face, construct, and criterion validity. However, in common with other recent trust and distrust scales (Cho, 2006; D. H. McKnight & Choudhury, 2006; D. H. McKnight et al., 2004; Weitzl, 2016), these subscales appear to have low discriminant validity, overly high internal consistency, and interitem correlations. That these issues are consistent between parts of the current scale and other, conceptually similar scales implies the CDISCS is equally ready for deployment. Still, additional administrations are warranted to

confirm the scale's generalizability to the non-Mechanical Turk population. I describe in this section the current state of each subscale in terms of reliability, validity, and readiness for deployments (summarized in Table 16), followed by a discussion of the limitations of the current work and future directions that would provide further validation for the current scale.

Disposition to Distrust Subscale

The Disposition subscale was developed by drawing from a rich body of theoretical and empirical work on the propensity to distrust as an individual trait, and several subscale items were adapted from existing distrust scales. This subscale differs from prior scales, as the Disposition subscale is meant to quantify an individual's general disposition rather than obtain a clinical diagnosis (e.g., Cattell & Mead, 2008). In addition, unlike other recent dispositional scales (D. H. McKnight et al., 2004) it features a response scale that does not rely on agreement/disagreement statements. This approach was chosen to reduce a common threat to survey validity, acquiescence bias: the tendency to prefer agreeing with opinion statements rather than disagree with them regardless of the respondent's true attitude (Knowles & Nathan, 1997).

This subscale appears to be, with one notable exception, highly reliable. Item-level retest correlations for the 6 items in this subscale were high across two deployments two weeks apart, *r*s > .67, and higher for the subscale as a whole, r = .92. The subscale features high internal consistency, Cronbach's $\alpha = .88$, and the average interitem correlation of .56 suggests that the items are sufficiently different

from each other to capture diverse aspects of the distrust disposition construct. Nevertheless, there are differences in disposition to distrust based on demographic differences in the Study 2 sample. There were no observable differences based on gender, but respondents who were younger, belonged to an ethnic minority group, or had some or no college scored higher on the disposition to distrust scale than those were older, identified as White or European-American, or had a college degree. This may reflect population-level differences in this personality trait, but it is possible that the scale itself requires centering based on respondent characteristics.

The Disposition subscale also appears to be valid by the criteria measured in this study. The item creation and reduction processes were theoretically-driven, to develop and retain items that fit two subconstructs that prior work described. These were (1) skepticism towards the motives of others, and (2) the belief that regardless of others' intent, it is strategically better to behave as though others will let them down. Items assessing each of these are in the final subscale, giving it good construct validity. The subscale also has high convergent validity, evidenced by the high correlation between this subscale and a prior validated measure of disposition to distrust, the 16 Personality Factors Mistrust scale. It also has high discriminant validity, as the correlations were low between this subscale, the three others from the current scale, and the brand trust and satisfaction scales included in Study 2.

However, data on the long-term reliability and criterion validity of this subscale are still needed. As a dispositional trait, Dispositional subscale scores should remain consistent over long periods of time. Additional deployments assessing the

same respondent over the course of months or years would allow a more confident assessment of whether this subscale is assessing a stable personality trait or a more transient state that changes over short periods of time. Finally, without behavioral data to suggest respondents with higher Disposition scores exhibit more distrustful behavior, the criterion validity for this subscale remains an open question. Studies correlating scores on this subscale to behavioral measures of distrust could be useful in demonstrating how effectively this subscale is measuring not only an internal state but one that also affects external behavior.

Incompetence, Malevolence, and Intentions Subscales

Although the remaining three subscales have different numbers of items in each (Incompetence: 4 items, Malevolence: 7 items, Intentions: 6 items), they currently perform similarly in terms of both reliability and validity. Although individual items in each subscale appear to be reliable (Study 1), and the subscales appear to adequately measure change due to an authentic intervention (Study 3), specific issues with discriminant validity and internal consistency (Study 2) warrant caution in considering the full scale ready for deployment.

In terms of reliability, each of these subscales features high item- and subscale-level retest reliability across two administrations two weeks apart. The subscales also appear to, with few exceptions, perform similarly across the demographic categories analyzed in Study 2. All three subscales had similar means and standard deviations across the ethnicity and education groups analyzed. Where the subscales differed was with respect to age and gender. Incompetence and

Intentions subscale scores appeared similar regardless of age or gender, but Malevolence scores were slightly negatively correlated with age, and men rated companies as about a half point more malevolent than women did.

However, I note that these age and gender differences in Malevolence scores are in the same direction and similar size to those for the Disposition subscale. Although only correlational, this provides some evidence that within the definitional framework McKnight and Chervany set up, dispositional differences in distrust may be more related to attributions of malintent and dishonesty to companies rather than attribution of incompetence. Consider also that the intervention in Study 3 more strongly affected perception of incompetence than malevolence, despite the news article describing an ethical failure rather than, say, a poor-quality product. Although Disposition is similarly correlated to both Incompetence and Malevolence in the Study 2 deployment, this is worth further investigation. It is possible that companies might expect less change in public perception of their malevolence than of their incompetence, and that difference could have implications for campaigns that seek to change public opinion about a company.

Study 3 provided encouraging criterion validity evidence regarding these three subscales. The intervention study provides a paired comparison of known-groups: first, a set of respondents who currently have an account with the bank in question should be expected to show lower than average distrust for that bank due to the ongoing nature of a business-to-consumer relationship.That was supported by the lower scores of each subscale for respondents in Study 3 compared to the general

subscale means observed in Study 2. Second, I provided a distrust-enhancing intervention to the participants which significantly increased their scores on all three subscales from their pre-test scores. This is evidence that even on a short timescale, each of these subscales is capable of measuring change in attributions of incompetence and malevolence to a company, as well as change in the intention to avoid that company based on new information. This indicates that the subscales are currently capable of assessing change in distrustful attitudes, beliefs, and intentions towards a company that they have a relationship with.

Implications for Theory

This study drew heavily on three models of trust and distrust: Komiak and Benbasat's dual-process model of trust and distrust, Lankton and McKnight's (2011) model of inter-entity/technological trust, and McKnight and Chervany's (2001) definitional model of distrust. Although the goal of the current thesis was to create a methodological tool rather than put these theories to the test, I share an insight from the current study that feeds back particularly into McKnight and Chervany's definitional model.

I return to the institutional features construct of McKnight and Chervany's model. During Study 1, the non-expert cognitive pretesting participants felt that word of mouth and prior experience were the institutional features that related most strongly to distrust. Contrary to McKnight and Chervany's model, lack of structural assurance did not appear to be relevant to their distrust for the companies they rated. A potential explanation for this discrepancy is that in their model, this construct is an

inverted trust construct, the opposite of their trust model's presence of structural assurance. Lacking structural assurance may be associated with lower trust, but I suggest that evidence of wrongdoing should theoretically have a greater direct effect on distrust. Study 3 provided evidence to that effect within the context of this survey, but future research could explore the effects of both positive and negative experiences and reputational information on evaluations of companies' distrust, as well as how individual differences in disposition affect the interpretation of institutional features that affect distrust.

Limitations and Future Directions

Currently, the greatest issue facing the three subscales is their overly high internal consistency. Up to a point, high internal consistency is evidence of construct validity, reflecting that each item in a scale taps into the same construct. I used Cronbach's alpha as the primary measure of internal consistency in the current thesis, for which a value of $\alpha > .8$ reflects a conservative standard of adequate internal consistency for scales with at least three items (Clark & Watson, 1995), At the same time, internal consistency that is too high ($\alpha > .90$) can result from having items that too unidimensionally assess the construct (Streiner, 2003). Such high internal consistency despite having items that purportedly measure different aspects of the same construct brings into question whether the items overlap too much, pointing to a lack of conceptual coverage in the subscales. However, this degree of high internal consistency is not unique to the current scale. Many scales assessing trust and distrust that are deployed in research have subscales with similarly high internal consistency

(Cho, 2006; D. H. McKnight & Choudhury, 2006; D. H. McKnight et al., 2004; Weitzl, 2016). Researchers and companies should therefore consider the CDISCS scale equally ready for deployment in this respect.

Two potential resolutions to this issue are suggested in prior literature (Streiner, 2003): the removal of redundant items , or alternatively, the addition of items that tap into other aspects of the construct. Items that were noticeably redundant were already removed prior to Study 2, but inter-item correlations were still high even between items that ostensibly measure different aspects of these constructs. For example, the items MAL1 "[Company]'s values seem (not at all different from my values…extremely different from my values)" and MAL5 "The company image that [Company] presents to the public is (not at all fake…extremely fake)" ask about different aspects of a company: the first, difference between the company's and the respondent's own values; the second, how dishonestly the companies portrays itself publicly. Despite being conceptually distinct, these items are highly positively correlated at $r = .79$, meaning 62% of the variance in one can be predicted by knowing the other. Similarly high inter-item correlations are found throughout the subscales. Again, this is not atypical within trust and distrust scales, but it is worth noting here as an area for potential improvement.

In addition to the high inter-item correlations within each subscale, the correlations between these three subscales are also quite high (Incompetence x Malevolence: .81, Malevolence x Intentions: .76, Incompetence x Intentions: .79). This is evidence of low discriminant validity based on the administration in Study 2.

Although these correlations appear similar, they have different theoretical implications. I begin by noting that as with the internal consistency issue, low discriminant validity is a common issue with trust and distrust scales. Cho's (2006) brand competence and benevolence subscale scores were correlated at over $r = .70$, and McKnight and Choudhury's (2006) distrust subscale scores were also highly intercorrelated. A recent study showed that two out of three brand trust scales frequently used in marketing research lacked discriminant validity with regards to reliability, safety, and honesty (Brudvig, 2015). These dimensions are conceptually related to the Incompetence and Malevolence subscales in the current scale, so these subscales' high intercorrelation has some precedent. This does not, however, mean that discriminant validity is not worth striving for.

Features of the survey deployments in the current study may have also contributed to the low discriminant validity and high correlations between items and subscales. One such issue is that I used company elicitation prompts to allow each respondent to name the company they would rate. The prompts in Study 1 were polarized, asking the participant to name the company which they had the most positive or most negative interactions with, or to name a company they used in the past but no longer use. This strategy was effective in garnering a wide set of companies and reducing the chance of respondents choosing to rate companies they have never interacted with. However, analyzing the open-ended responses reveals that few, if any, respondents chose companies they felt neutrally about, likely due to the

prompt eliciting the company who the respondent felt most saliently and (likely) strongly about.

In Study 2, I attempted to reduce the impact of this salience effect by replacing and supplementing the prompts with additional, less polarized prompts. It is, however, possible that the use of company elicitation prompts rather than specifying a company may have limited the companies chosen by participants to those they felt most strongly positively or negatively about. If that is the case, additional deployments that ask all participants to rate a single company with known characteristics may provide useful evidence as to the general performance of the subscales. For example, a company that is widely believed to have excellent products but be quite malevolent could be assessed to see whether respondents respond oppositely to the Incompetence and Malevolence subscales. If such a deployment showed lower inter-scale correlations, that would support the discriminant validity of the scale. One concern with such a deployment is that it may require a high level of familiarity with the target company. Less familiar respondents may have a general sense of the company but lack sufficient experience to meaningfully rate different aspects of the company. As in the present work, it would be necessary to recruit respondents who are sufficiently familiar with the company they are evaluating, and to directly ask respondents to rate their familiarity with the target company.

Another concern is the use of Mechanical Turk as the sole source to recruit participants. Several researchers have demonstrated that the Turk population differs from the general US population in several respects, including demographics and

attention (Berinsky, Huber, & Lenz, 2012). I attempted to address these concerns in the current study by analyzing for demographic differences in the subscales and by compensating participants at a rate that was higher than typical for Turk: $10 per hour, matching California's current minimum hourly wage. I chose to do so both for ethical reasons as well as to minimize the risk of respondents satisficing due to a need to maximize their earnings on the site. Although the study received universally high ratings on Turkopticon, a site that allows Turk workers to rate requesters after completing a task for them, I acknowledge that the high wage paid to participants may have simply been effective at recruitment rather than at minimizing individual respondents' inattention or speeding through the scale. I believe that additional deployments such as panel studies that more closely resemble natural deployments to the general population would be useful in determining how well the subscales perform with a non-Turk sample.

Considering the similarities in internal consistency and discriminant validity concerns between the CDISCS and prior trust and distrust scales, I must mention that trust and distrust research often attempts to tease apart distinctions in subconstructs that are only rarely or implicitly considered by most consumers. Although pretesting in Study 1 showed that a set of experts and non-experts believed the items in each subscale for the most part included ideas they had spontaneously thought about prior to reading the scale, there were high, negative correlations between these subscales and the brand satisfaction scale I deployed in Study 2. It makes logical sense that brand satisfaction would be negatively associated with distrust, but the high

correlations may indicate that even items intended and written to tap into specific aspects of distrust, in fact, simply reflect a lack of satisfaction with the company. This is a topic worth exploring further, but even a strong negative relationship between brand distrust and brand satisfaction does not invalidate the need for either distrust or dissatisfaction measures. As discussed in the literature review of this thesis, distrust is not the mirror image of trust and requires measures that assess it separately from trust. The current study provides just that, validating the CDISCS in hopes that it becomes a useful tool for future research to, among other things, more deeply explore the relationship between trust and distrust.

Conclusion

This thesis details the creation and validation of a new instrument that measures an individual's distrust in a specific company, the CDISCS scale. This scale fills a methodological gap left by prior work on trust and distrust: the scale is flexible across companies; is theory-driven, incorporating aspects of multiple theoretical and empirical models of distrust; and is modular, with four subscales that can be deployed individually or together to assess different aspects of distrustful disposition, beliefs and intentions. In particular, I highlight the scale's high criterion validity: Study 3 demonstrated the CDISCS's ability to measure change in consumer attitudes, beliefs, and intentions after an actual scandal. The extant concerns with this scale regarding discriminant validity and internal consistency are comparable to those found in other scales that assess related constructs, implying they may reflect features of the underlying constructs rather than deficiencies in the current scale. Further, these

concerns may be reduced as additional deployments are conducted with other populations and target companies. Continued iteration on the scale should only improve its ability to capture the key dimensions of distrust in an easily deployable format. I recommend that managers, marketers, and researchers interested in tracking consumers' distrust of specific companies over time consider deploying the CDISCS scale.

References

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514.

Acquisti, A., & Grossklags, J. (2004). Privacy attitudes and privacy behavior. In *Economics of Information Security* (pp. 165–178). Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/1-4020-8090-5_13.pdf

Afzal, H., Khan, M. A., Rehman, K. ur, Ali, I., & Wajahat, S. (2009). Consumer's trust in the brand: Can it be built through brand reputation, brand competence and brand predictability. *International Business Research*, *3*(1). https://doi.org/10.5539/ibr.v3n1p43

Baier, A. (1986). Trust and Antitrust. *Ethics*, *96*(2), 231–260. https://doi.org/10.2307/2381376

Benamati, J. "Skip," Serva, M. A., & Fuller, M. A. (2010). The Productive Tension of Trust and Distrust: The Coexistence and Relative Role of Trust and Distrust in Online Banking. *Journal of Organizational Computing and Electronic Commerce*, *20*(4), 328–346. https://doi.org/10.1080/10919392.2010.516632

Benedicktus, R. L., Brady, M. K., Darke, P. R., & Voorhees, C. M. (2010). Conveying trustworthiness to online consumers: Reactions to consensus, physical store presence, brand familiarity, and generalized suspicion. *Journal of Retailing*, *86*(4), 322–335. https://doi.org/10.1016/j.jretai.2010.04.002

Blaszczynski, A., Dumlao, V., & Lange, M. (1997). "How much do you spend gambling?" Ambiguities in survey questionnaire items. *Journal of Gambling Studies*, *13*(3), 237–252.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Cain, N. M., Lukowitsky, M. R., & Wright, A. G. C. (2014). Multidimensional Personality Questionnaire. In *The Encyclopedia of Clinical Psychology*. John Wiley & Sons, Inc. Retrieved from http://dx.doi.org/10.1002/9781118625392.wbecp370

Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web survey methodology*. Sage.

Cattell, H. E., & Mead, A. D. (2008). The sixteen personality factor questionnaire (16PF). *The SAGE Handbook of Personality Theory and Assessment, 2*, 135–178.

Chang, Y.-S., Fang, S.-R., & others. (2013). Antecedents and distinctions between online trust and distrust: Predicting high-and low-risk Internet behaviors. *Journal of Electronic Commerce Research, 14*(2), 149–166.

Chen, P., Dhanasobhon, S., & Smith, M. (2001). All reviews are not created equal: The disaggregate impact of reviews and reviewers at Amazon.com. *Heinz College Research*. Retrieved from http://repository.cmu.edu/heinzworks/55

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

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method* (4th edition). Hoboken: Wiley.

Egelman, S., & Peer, E. (2015). Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS) (pp. 2873–2882). ACM Press. https://doi.org/10.1145/2702123.2702249

Ericsson, K. A., & Simon, H. A. (1998). How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind, Culture, and Activity*, *5*(3), 178–186. https://doi.org/10.1207/s15327884mca0503_3

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, *30*(2), 407–429.

Greenlaw, C., & Brown-Welty, S. (2009). A Comparison of Web-Based and Paper-Based Survey Methods: Testing Assumptions of Survey Mode and Response Cost. *Evaluation Review*, *33*(5), 464–480. https://doi.org/10.1177/0193841X09340214

Ha, H. (2004). Factors influencing consumer perceptions of brand trust online. *Journal of Product & Brand Management*, *13*(5), 329–342. https://doi.org/10.1108/10610420410554412

Holden, C. J., Dennie, T., & Hicks, A. D. (2013). Assessing the reliability of the M5-120 on Amazon's mechanical Turk. *Computers in Human Behavior*, *29*(4), 1749–1754. https://doi.org/10.1016/j.chb.2013.02.020

Internet-Based Research: Online Surveys. (2015). Retrieved May 25, 2016, from https://irb.ucsc.edu/customer/en/portal/articles/1729145-internet-based-research-online-surveys?b_id=2829

Janiszewski, C., & Meyvis, T. (2001). Effects of Brand Logo Complexity, Repetition, and Spacing on Processing Fluency and Judgment. *Journal of Consumer Research*, *28*(1), 18–32. https://doi.org/10.1086/321945

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus: Privacy calculus: dispositions and affect. *Information Systems Journal*, *25*(6), 607–635. https://doi.org/10.1111/isj.12062

Kleiman, T., Sher, N., Elster, A., & Mayo, R. (2015). Accessibility is a matter of trust: Dispositional and contextual distrust blocks accessibility effects. *Cognition*, *142*, 333–344. https://doi.org/10.1016/j.cognition.2015.06.001

Komiak, S. Y. X., & Benbasat, I. (2008). A two-process view of trust and distrust building in recommendation agents: A process-tracing study. *Journal of the Association for Information Systems*, *9*(12), 727–747.

Lankton, N. K., & McKnight, D. H. (2011). What does it mean to trust Facebook?: examining technology and interpersonal trust beliefs. *ACM SIGMIS Database*, *42*(2), 32–54.

Larzelere, R. E., & Huston, T. L. (1980). The Dyadic Trust Scale: Toward Understanding Interpersonal Trust in Close Relationships. *Journal of Marriage and the Family*, *42*(3), 595. https://doi.org/10.2307/351903

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust And Distrust: New Relationships and Realities. *Academy of Management Review*, *23*(3), 438–458. https://doi.org/10.5465/AMR.1998.926620

Li, Y. (2012). Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, *54*(1), 471–481. https://doi.org/10.1016/j.dss.2012.06.010

Lowry, P. B., Schuetzler, R. M., Giboney, J. S., & Gregory, T. A. (2015). Is Trust Always Better than Distrust? The Potential Value of Distrust in Newer Virtual Teams Engaged in Short-Term Decision-Making. *Group Decision and Negotiation*, *24*(4), 723–752. https://doi.org/10.1007/s10726-014-9410-x

Lowry, P., Vance, A., Moody, G., Beckman, B., & Read, A. (2008). Explaining and predicting the impact of branding alliances and web site quality on initial

consumer trust of e-commerce web sites. *Journal of Management Information Systems*, *24*(4), 199–224. https://doi.org/10.2753/MIS0742-1222240408

Marsden, P. V., & Wright, J. D. (Eds.). (2010). *Handbook of survey research* (2. ed). Bingley: Emerald.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. https://doi.org/10.3758/s13428-011-0124-6

McKnight, D. H., & Chervany, N. L. (2001). Trust and distrust definitions: One bite at a time. In *Trust in Cyber-societies* (pp. 27–54). Springer. Retrieved from http://link.springer.com/10.1007/3-540-45547-7_3

McKnight, D. H., & Choudhury, V. (2006). Distrust and trust in B2C e-commerce: Do they differ? In Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet (pp. 482–491). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1151527

Montague, E. N. H., Kleiner, B. M., & Winchester, W. W. (2009). Empirically understanding trust in medical technology. *International Journal of Industrial Ergonomics*, *39*(4), 628–634. https://doi.org/10.1016/j.ergon.2009.01.004

Moody, G. D., Galletta, D. F., & Lowry, P. B. (2014). When trust and distrust collide online: The engenderment and role of consumer ambivalence in online consumer behavior. *Electronic Commerce Research and Applications*, *13*(4), 266–282. https://doi.org/10.1016/j.elerap.2014.05.001

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x

Pasek, J., & Krosnick, J. A. (2010). Optimizing survey questionnaire design in political science: insights from psychology. *Oxford Handbook of American Elections and Political Behavior*, 27–50.

Peer, E., Paolacci, G., Chandler, J., & Mueller, P. (2012). Selectively recruiting participants from Amazon Mechanical Turk using qualtrics. *Available at SSRN 2100631*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2100631

Pettit, P. (2008). Trust, reliance, and the internet. In M. J. van den Joven & J. Weckert (Eds.), *Information Technology and Moral Philosophy* (p. 161). Cambridge University Press.

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet Versus Mailed Questionnaires: A Randomized Comparison. *Journal of Medical Internet Research*, *6*(3), e29. https://doi.org/10.2196/jmir.6.3.e29

Rohrmann, B. (2007). Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data. *Project Report. University of Melbourne, Melbourne*. Retrieved from http://www.rohrmannresearch.net/pdfs/rohrmann-vqs-report.pdf

Rose, A., Peters, N., Shea, J. A., & Armstrong, K. (2004). Development and testing of the health care system distrust scale. *Journal of General Internal Medicine, 19*(1), 57–63.

Rouse, S. V. (2015). A reliability analysis of Mechanical Turk data. *Computers in Human Behavior, 43,* 304–307. https://doi.org/10.1016/j.chb.2014.11.004

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement, 9*(4), 367–373.

Sebastiani, G., Tinto, A., Battisti, A., & De Palma, E. (2013). Cognitive Interviewing as a Tool for Improving Data Quality in Surveys: Experiences in Istat. Retrieved from http://new.sis-statistica.org/wp-content/uploads/2013/09/RS10-Cognitive-Interviewing-as-a-Tool-for-Improving-Data-Quality-in-Surveys-Experiences-in-Istat.pdf

Shea, J. A., Micco, E., Dean, L. T., McMurphy, S., Schwartz, J. S., & Armstrong, K. (2008). Development of a Revised Health Care System Distrust Scale. *Journal of General Internal Medicine, 23*(6), 727–732. https://doi.org/10.1007/s11606-008-0575-3

Simms, L. J., Goldberg, L. R., Roberts, J. E., Watson, D., Welte, J., & Rotterman, J. H. (2011). Computerized Adaptive Assessment of Personality Disorder: Introducing the CAT–PD Project. *Journal of Personality Assessment, 93*(4), 380–389. https://doi.org/10.1080/00223891.2011.577475

Simons, D. J., & Chabris, C. F. (2012). Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. *PLoS ONE, 7*(12), e51876. https://doi.org/10.1371/journal.pone.0051876

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. https://doi.org/10.1016/j.tourman.2010.12.011

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659–677.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed). Boston: Pearson/Allyn & Bacon.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS ONE*, *8*(9), e73990. https://doi.org/10.1371/journal.pone.0073990

Willis, G. B. (2004). Cognitive Interviewing: A Tool for Improving Questionnaire Design: A Tool for Improving Questionnaire Design. Sage Publications.

Woods, C. M. (2006). Careless Responding to Reverse-Worded Items: Implications for Confirmatory Factor Analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186–191. https://doi.org/10.1007/s10862-005-9004-7

Yeager, D. S., & Krosnick, J. A. (2012). Does Mentioning "Some People" and "Other People" in an Opinion Question Improve Measurement Quality? *Public Opinion Quarterly*, *76*(1), 131–141. https://doi.org/10.1093/poq/nfr066

Appendix

Appendix 1. Full Scale based on Results from Study 2.

Disposition to distrust

DIS1. In general, I am...

1 Not at all suspicious of people I don't know well
2 Slightly suspicious of people I don't know well
3 Moderately suspicious of people I don't know well
4 Very suspicious of people I don't know well
5 Extremely suspicious of people I don't know well

DIS2. I assume people that I meet are...

1 Not at all selfish
2 Slightly selfish
3 Somewhat selfish
4 Very selfish
5 Extremely selfish

DIS3. When I have to ask others for help, I...

1 Never think about whether they might let me down
2 Rarely think about whether they might let me down
3 Sometimes think about whether they might let me down
4 Usually think about whether they might let me down
5 Always think about whether they might let me down

DIS4. Most people...

1 Never lie when they would benefit from doing so
2 Rarely lie when they would benefit by doing so
3 Sometimes lie when they would benefit by doing so
4 Usually lie when they would benefit by doing so
5 Always lie when they would benefit by doing so

DIS5. People are better off assuming that others will...

1 Never let them down
2 Rarely let them down
3 Sometimes let them down
4 Usually let them down
5 Always let them down

DIS6. I assume that promises will...

1 Never be broken
2 Rarely be broken
3 Sometimes be broken
4 Usually be broken
5 Always be broken

Incompetence

INC1. [Company] is...

1 Not at all incompetent
2 Slightly incompetent
3 Moderately incompetent
4 Very incompetent
5 Extremely incompetent

INC2. How skeptical are you about whether [Company] is capable of doing a good job?

1 Not at all skeptical
2 Slightly skeptical
3 Moderately skeptical
4 Very skeptical
5 Extremely skeptical

INC3. To what extent does [Company] fall short at what it tries to do?

1 Not at all short
2 Slightly short
3 Moderately short
4 Very short
5 Extremely short

INC4. [Company]'s policies for fixing issues that customers or users encounter are...

1 Not at all inadequate
2 Slightly inadequate
3 Moderately inadequate
4 Very inadequate
5 Extremely inadequate

Malevolence

MAL1. [Company]’s values seem...

1 The same as my values
2 Slightly different from my values
3 Moderately different from my values
4 Very different from my values
5 Extremely different from my values

MAL2. [Company]...

1 Never tries to take advantage of people
2 Rarely tries to take advantage of people
3 Sometimes tries to take advantage of people
4 Usually tries to take advantage of people
5 Always tries to take advantage of people

MAL3. I believe that [Company]...

1 Never acts immorally
2 Rarely acts immorally
3 Sometimes acts immorally
4 Usually acts immorally

5 Always acts immorally

MAL4. I believe that [Company]...

1 Never tries to trick people
2 Rarely tries to trick people
3 Sometimes tries to trick people
4 Usually tries to trick people
5 Always tries to trick people

MAL5. The company image that [Company] presents to the public is...

1 Not at all fake
2 Slightly fake
3 Moderately fake
4 Very fake
5 Extremely fake

MAL6. Altogether, the policies and laws that protect people from harm by [Company] are...

1 Not at all lacking
2 Slightly lacking
3 Moderately lacking
4 Very lacking
5 Extremely lacking

MAL7. The stories I have heard about [Company] as a company are...

1 Never negative
2 Rarely negative
3 Sometimes negative
4 Usually negative
5 Always negative

Intentions

INT1. In the future, I will...

1 Certainly not avoid [Company]
2 Probably not avoid [Company]
3 Possibly avoid [Company]
4 Quite likely avoid [Company]
5 Certainly avoid [Company]

INT2. How strongly do you feel about avoiding [Company] in the future?

1 Not at all strongly
2 Slightly strongly
3 Moderately strongly
4 Very strongly
5 Extremely strongly

INT3. If you wanted something that only [Company] could provide, would you use [Company] or do without?

1 Definitely use [Company]
2 Probably use [Company]

3 About 50:50 to use or do without
4 Probably do without
5 Definitely do without

INT4. If [Company]'s products or services were better than their competitors, how likely is it that you would go with a competitor over [Company]?

1 I would certainly not go with a competitor
2 I would be unlikely to go with a competitor
3 I would perhaps go with a competitor
4 I would quite likely go with a competitor
5 I would certainly go with a competitor

INT5. If competitors' products or services were more convenient than [Company]'s, how likely would you be to use a competitor instead of [Company]?

1 I would certainly not use a competitor
2 I would be unlikely to use a competitor
3 I would perhaps use a competitor
4 I would quite likely use a competitor
5 I would certainly use a competitor

INT6. If avoiding [Company]'s products or services meant you would miss out on something you cared about, would you avoid or use their products or services?

1 Definitely use [Company]
2 Probably use [Company]
3 About 50:50 to avoid or use [Company]
4 Probably avoid [Company]
5 Definitely avoid [Company]