[Inner endpaper]

HEC MONTRÉAL
École affiliée à l'Université de Montréal

Three Essays on Choice Overload, Recommendation Systems and Failures/Transgressions by AI-based Recommendation Systems

par
Anshu Pradeep Suri

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(option Marketing)

Août 2021

HEC MONTRÉAL
École affiliée à l'Université de Montréal

Cette thèse intitulée :

Three Essays on Choice Overload, Recommendation Systems and Failures/Transgressions by AI-based Recommendation Systems

Présentée par :

Anshu Pradeep Suri

a été évaluée par un jury composé des personnes suivantes :

Jonathan Deschenes
HEC Montréal
Président-rapporteur

Sylvain Sénécal
HEC Montréal
Directeur de recherche

Yany Grégoire
HEC Montréal
Codirecteur de recherche

Caroline Roux
Concordia University
Membre du jury

Mansur Khamitov
Indiana University
Examinateur externe

Guy Paré
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Cette thèse se compose de trois essais. Le premier essai explore le concept de surcharge de choix de produits (*choice overload*). Étant donné que des recherches antérieures ont noté des effets négatifs et positifs liés au fait de faire un choix de produits parmi une vaste gamme de produits, nous explorons ce concept de surcharge de choix à l'aide de mesures psychophysiologiques obtenues et analysées via un logiciel de reconnaissance faciale automatisé (FaceReader). Nous montrons que les consommateurs ressentent des émotions mixtes - en particulier, du bonheur et de la colère - tout en choisissant parmi des ensembles de choix plus larges. Une enquête plus approfondie sur le processus sous-jacent indique que la difficulté de choisir médie l'effet d'un assortiment de produits plus grand (vs. plus petit) sur les émotions négatives, et que le besoin de cognition modère l'effet des choix accrus sur les émotions positives.

Le deuxième essai utilise un nouveau système de recommandation adaptatif, qui examine si les recommandations dynamiques - qui apparaissent lorsque les consommateurs rencontrent une difficulté de choix plus élevée - aident ou entravent le processus de prise de décision. Un courant de recherche dicte que de telles recommandations dynamiques faciliteraient la tâche de prise de décision des consommateurs en réduisant l'ensemble de considération. D'un autre côté, un deuxième courant de recherche indique que de telles recommandations adaptatives qui sont présentées au milieu de la prise de décision du consommateur pourraient au contraire interférer davantage avec le processus de prise de décision et accroître la difficulté de choix en élargissant l'ensemble de considération. Construit à l'aide d'outils neurophysiologiques (EEG), ce nouveau système indique que de tels systèmes de recommandations dynamiques peuvent en effet aider les consommateurs en diminuant leur charge cognitive.

Enfin, le troisième essai examine les réponses des consommateurs lorsque des systèmes d‘intelligence artificielle (IA), tels que le système de recommandation, échouent. Étant donné que les consommateurs adoptent de plus en plus de tels agents d'IA dans tous les domaines de la vie, qu'il s'agisse de les aider à prendre des décisions d'achat ou d'éteindre les lumières de leur salon, les consommateurs dépendent plus que jamais de l'IA. Cet essai explore comment l'anthropomorphisme de tels systèmes d'IA conduit à des réactions positive ou négatives des consommateurs après un échec. Plus précisément, en introduisant un modérateur important dans ce contexte – le type d'échec – nous montrons que pour les échecs de performance, les systèmes anthropomorphiques (vs. non anthropomorphiques) conduisent à des réponses négatives plus modérées de la part des consommateurs. D'un autre côté, pour les échecs de bienveillance, de tels systèmes anthropomorphiques (vs. non anthropomorphiques) suscitent des réactions plus négatives des consommateurs. À l'aide de quatre études, dont l'une examine les tweets des consommateurs dans le monde réel et une autre qui a utilisé un véritable échec de service à l'aide d'un robot conversationnel (chatbot) de recommandation de produits, cet essai démontre les résultats décrits ci-dessus. Cet essai examine également le processus sous-jacent derrière ces résultats.

Mots clés : surcharge de choix, paradoxe du choix, émotions mixtes, système de recommandation, recommandations dynamiques, échec de l'IA, échec de la recommandation, anthropomorphisme, mesures psychophysiologiques, EEG

Méthodes de recherche : Conception expérimentale, mesures psychophysiologiques, analyses des sentiments, mesures neurophysiologiques

Abstract

This thesis consists of three essays. The first essay explores the concept of choice overload. Given that past research has noted adverse, as well as beneficial consumer outcomes due to increased choice, we explore this concept of choice overload using psychophysiological measures obtained and analyzed through an automated facial recognition software (FaceReader). Presenting a novel finding, we show that consumers experience increased mixed emotions – specifically, increased happiness and anger – while selecting through larger choice sets. Further investigation of the underlying process indicates that while choice difficulty mediates the effect of larger (vs. smaller) choice set on negative emotions, need for cognition moderates the effect of increased choices on positive emotions.

The second essay employs a novel adaptive recommendation system, that examines if dynamic recommendations – which appear when consumers experience a higher choice difficulty – help or hinder the decision-making process. One stream of research suggests that such dynamic recommendations would ease consumers' decision-making task by reducing the consideration set. On the other hand, a second stream of research indicates that such adaptive recommendations that are presented in the midst of consumer decision-making could instead interfere with the decision-making process and further enhance choice difficulty by widening the consideration set. Built using neurophysiological tools (EEG), this novel system indicates that such dynamic recommendation systems may indeed help consumers by decreasing their cognitive load.

Finally, the third essay examines consumer responses when artificially intelligent (AI) systems, such as recommendation system fail consumers. Given that consumers are increasing adopting such AI-agents in every field of life, from helping them make purchase decisions to switching off the lights in their living room, consumers are depending on AI more than ever. This essay explores how anthropomorphism of such AI systems leads to better and worse consumer reactions after a failure. Specifically, introducing an important moderator in this context – the type of failure – we show that for performance failures, anthropomorphic (vs. non- anthropomorphic) systems lead to milder negative responses from consumers. On the other hand, for benevolence failures, such anthropomorphic (vs. non-anthropomorphic) systems elicit worsened consumer responses. Using four studies, one of which examines real-world consumer tweets, and another which employed a real service failure (and an automated analysis of facial expressions) using a recommendation chatbot, this essay demonstrates the findings outlined above. This essay also investigates the underlying process behind these differential outcomes.

Keywords: choice overload, paradox of choice, mixed emotions, recommendation system, dynamic recommendations, AI failure, recommendation failure, anthropomorphism, psychophysiological measures, EEG

Research methods: Experimental design, psychophysiological measures, sentiment analyses, neurophysiological measures

Table of Contents

Contents

List of Tables and Figures

Acknowledgements

The saying, "It takes a village to get a PhD" could not be any truer for me. This journey would not have been possible without the numerous people who have supported me through out. As I list out all the people whom I am thankful to, my supervisors of course top the list. I would like to thank my supervisor Sylvain Sénécal for making this PhD a reality for me. From admitting me to the PhD program to encouraging me after my disastrous job interviews, words will never be enough to express how thankful I am to you for everything. You have spent countless hours on improving me as a researcher and helping me manage my PhD life in general; you have shown trust in me at times when I did not trust myself, and I will always be thankful for your generosity. I would like to thank my co-supervisor Yany Grégoire for guiding me on research, teaching and life in general. Yany, you have always pushed me out of my comfort zone to be a better researcher. From including me on your research projects as a first year PhD student, to advising me life issues in general, you have always looked out for me, and I very thankful for that. You two are the kindest mentors one could hope for, and words will never suffice to express my gratitude.

My note of gratitude would be incomplete without thanking Caroline Roux, who has been extremely selfless and generous in helping me succeed as an academic. Caroline, I have learned so much from and because of you, and I hope I can give back selflessly to others, as you have to me. I would also like to thank Ali Tezer, who has made me a better

researcher, and has guided me on every step of the way. From refining my research ideas to my job market preparation, Ali is one of the most generous academics that I've met. My list would also be incomplete without thanking Jonathan Deschênes, who has worked tirelessly for the PhD students at HEC Montreal. I would also like to thank the rest of the marketing faculty at HEC Montreal for helping and guiding me throughout this journey.

I would also like to thank the everyone at the Tech3lab, especially Salima Tazi and Edouard Freve-Guerin for making my research (and life) easier.

This journey would not have been possible without my colleagues in the PhD program. I would especially like to thank Bo Huang (whom I jokingly address as my Asian brother), who has been my friend, family and the best research partner. Bo, without you, my life would have been miserable, thank you for being my family in Montreal. I would also like to thank my work family at Chaire de commerce Omer DeSerres: Hamid Shaker, Kamran Moazeni and William Blais, you guys made this journey much more enjoyable than it would have been.

This list also remains incomplete without thanking my mentor and friend Marie-Louise Radanielina-Hita. ML, you have guided my way through teaching, research, and life in general, thank you for being the incredibly generous woman that you are.

Lastly, I would like to thank my family for making today a reality for me. I was the first woman in my family to move abroad for a career, and this wouldn't have been possible without my parents' support. From funding my expensive MBA in the US, to ignoring the societal rules in India, my parents did what was unimaginable for many in India. Finally, I would like to thank my husband, Shravan for being the best partner I could have asked for. Like I always say, you are a better spouse than I ever will be. From ruining your sleep by being a night owl and working on my research in the wee hours of the morning, to working through our vacation, I have done it all – and yet you choose to never complain about it. Thank you for being the perfect partner that you are!

Introduction

Today's world is replete with choices. For instance, a simple search of laptops on Bestbuy.com results in more than 700 different choices. While traditional wisdom would suggest that an increase in the assortment size would result in improved customer outcomes, literature has often noted otherwise. This phenomenon, called choice overload – the existence of adverse consumer outcomes as a consequence of increased choice (Scheibehenne et al., 2010) – has received significant attention in the literature (see Chernev et al., 2015 for a meta-analysis). However, some research in this stream (Scheibehenne et al., 2009), including a popular meta-analyses with 995 citations (Scheibehenne et al., 2010) did not find the existence of this phenomenon, generating some debate in this context. Essay 1 contributes to the debate on choice overload, by using objective psychophysiological measures and analyzing a consumer outcome that has traditionally received less attention in this context: consumers' emotions when they make the purchase decision.

Using consumers' real-time emotions and an automated analysis of facial expressions in a task-based study wherein participants selected a laptop from the assortment (small vs. large), Essay 1 shows that an increase in assortment size leads to increased positive as well as negative emotions. Specifically, the automated analysis of facial emotions obtained via a FaceReader showed that consumer experience an increase in the discrete emotions of happiness, as well as anger, as the assortment size increases. These results indicate the existence of mixed emotions, as a function of assortment size. Essay 1 further shows that these two outcomes result due to two different underlying processes. While individuals with higher need for cognition experience an increase in positive emotions as a result of an increase in assortment size, negative emotions are a result of an increase in choice difficulty that results due to larger assortment size.

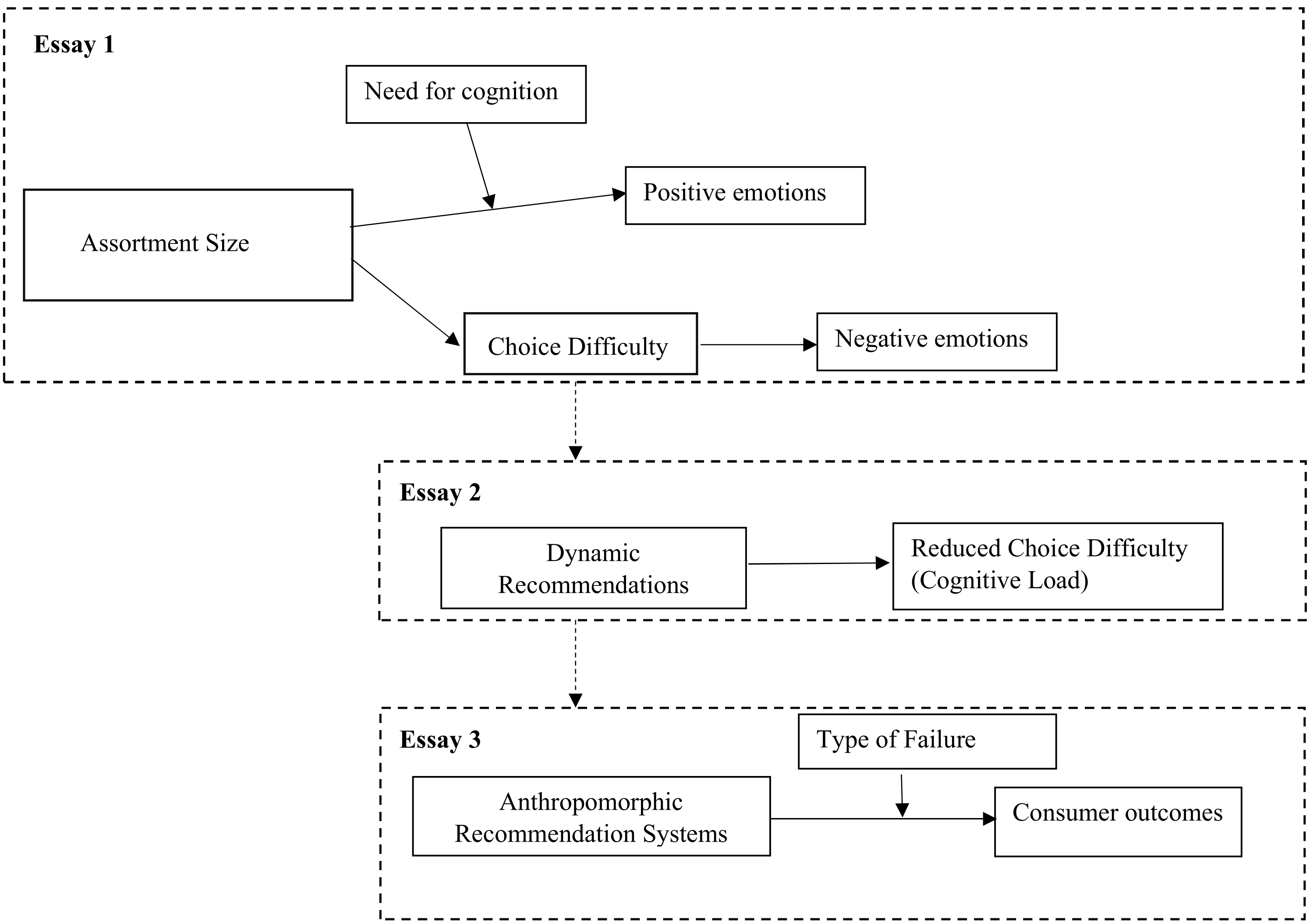


Figure 1. Overall Conceptual Framework of the Thesis

Given the findings of Essay 1, which indicate that consumers experience increased positive and negative emotions as the assortment size increases, Essay 2 investigates if the negative outcomes associated with an increase in choice can be subverted using a recommendation agent. In other words, considering that Essay 1 shows that consumers experience negative outcomes due to the increased choice difficulty that results with an enlargement of the assortment size, Essay 2 investigates if recommendation systems could ease this choice difficulty.

Contributing to a better understanding of adaptive personalized systems, this study employed and tested a novel, propriety recommendation system that uses neurophysiological measures to assess consumers' cognitive load in real-time and determine if the purchase decision leads to increased choice difficulty. Specifically using a laptop selection task, Essay 2 adapted and employed Threshold Reactive Adaptive Dynamic Spectrum (ThReADS), which categorized consumers' cognitive load (in real-time) into 5 categories: 0 (very low) to 4 (very high), as they progressed with the selection task. This personalized recommendation system is able to adapt the webpage to highlight product recommendations to consumers when the system assesses the cognitive load of consumers to be sufficiently high. However, such dynamic, adaptive recommendation systems, which highlight suitable options from within the initial set could potentially lead to two possible outcomes for the consumer: 1) Narrowing of the consideration set, and the alleviation of choice difficulty thereof or 2) Further enlargement of the consideration set and aggravation of choice difficulty as a result. Investigating these competing hypotheses through neurophysiological measures (using Electroencephalography (EEG)), the findings of Essay 2 indicate that such dynamic recommendation systems have the potential of reducing consumers' choice difficulty. This essay further investigates the framing implications for such context-aware, dynamic recommendations, to suggest that recommendations that are framed as those suggested by "similar other consumers" (vs. personalized for the user) lead to better consumers outcomes.

Building on Essay 2, which indicates that recommendation systems are indeed able to ease consumers' choice difficulty, Essay 3 investigates an important post-usage phenomenon in the context of recommendation systems: What if such systems fail consumers' expectations of completing the task optimally? Consumers are increasingly adopting task-based AI systems such as Amazon's Alexa and Google Home to accomplish everyday mundane tasks such as looking up suitable restaurants in the vicinity to browsing the latest musical releases. Further, given consumers' growing dependence upon such AI-agents, more and more businesses are anthropomorphizing such AI-agents using personification elements such as gendering the AI, naming it, or attributing it with humanlike physical attributes. Essay 3 investigates the repercussions of personification of AI-agents, when such agents fail. Using the context of recommendation systems in a number of studies, this essay investigates the consequences of failures that are attributed to anthropomorphic (vs. non- anthropomorphic) technological agents. Specifically, introducing an important moderator in the context of failures – the type of failure, this essay investigates the conditions wherein failures by anthropomorphic agents fare better versus worse. This essay establishes that while consumers exhibit attenuated negative reactions to performance failures committed by humanlike anthropomorphic (vs. non- anthropomorphic) systems, benevolence failures committed by such agents elicit increased negative responses. Essay 3 further investigates the underlying process for this phenomenon of differential outcomes for failures committed by anthropomorphic technological agents, highlighting contexts wherein anthropomorphizing the AI-agent buffers (vs. worsens) the consumer outcomes when such agents fail.

Chapter 1
Choice – A Tyranny or a Boon: A Physiological Assessment of Consumers' Emotions to "The Paradox of Choice"

Abstract

The notion of 'choice overload' has been debated in past research, given the contradictory findings regarding the adverse effects of "too many choices" being offered to today's consumers. Using psychophysiological measures, and an automatic detection of facial emotions, the current research presents a novel finding: more choice generates both greater negative and positive emotions in consumer. Further investigation indicates that while the effect of assortment size on negative emotions is mediated by difficulty in making a choice, the effect of choice overload on positive emotions is moderated by the level of the consumer's need for cognition. Thus, the current study not only helps reconcile the contradictory findings of past research on choice overload, but also sheds some light on the processes that result in the simultaneous existence of both negative and positive emotions that occur when choice variety is increased.

1.1 Introduction

Consider that Emma is at a supermarket, looking to purchase a bag of chips, and she comes across the aisle of chips which contains more than 30 different varieties. The question central to the current research is whether such a huge assortment of chips would generate positive or negative emotions in Emma. While traditional wisdom would dictate that a larger variety would lead to improved consumer outcomes, researchers have noted otherwise. For instance, in their seminal research, Iyengar and Lepper (2000) noted that while a huge assortment of jams in a supermarket attracted more consumers, it also led to fewer purchases. As contradictory as this finding may sound, a significant amount of research in the area of choice overload has noted similar findings with respect to different consumer outcomes such as choice deferral (Shah & Wolford, 2007), dissatisfaction (Chernev, 2006), and regret (Inbar et al., 2011).

The existence of choice overload – defined as “the adverse consequences due to an increase in the number of options to choose from” (Scheibehenne et al., 2010. p. 209) – has generated some debate. For instance, two meta-analyses on this topic (Chernev et al., 2015; Scheibehenne et al., 2010) note contradictory findings on the existence of this phenomenon. Analyzing 50 studies, Scheibehenne et al. (2010) found no effect of the number of options available to consumers on decision outcomes. On the other hand, an analysis of 53 studies by Chernev et al. (2015) showed that the effect of assortment size on consumer outcomes is moderated by four variables (choice set complexity, decision task difficulty, preference uncertainty, and decision goal) in this context.

We suggest that physiological measures might be able to shed some light on consumers' reactions to assortment sizes, contributing to a better understanding of this phenomenon. The current study thus contributes to the literature on choice overload by investigating it using physiological tools to assess consumers' reactions to varying assortment sizes. Further, while past research on choice overload has investigated several consumer outcomes such as choice deferral, satisfaction, regret, and confusion (e.g. Inbar et al., 2011; Iyengar & Lepper, 2000), scant attention has been paid to the emotions that consumers experience *as they make this decision*. A recent study (Tang et al., 2017) is one of the few research that examines the effect of choice overload on consumer emotions. However, the current research differs from the study conducted by Tang et al. (2017) in two important ways. First, as opposed to using self-report measures employed by Tang et al. (2017), we use objective psychophysiological measures to assess consumers' emotions. Prior research on choice overload has mainly investigated consumers' responses via self-report instruments (Diehl & Poynor, 2010; Tang et al., 2017) or by observing their actual behaviors (Iyengar & Lepper, 2000). While both these approaches of theory testing are essential for research, psychophysiology can complement these methods by providing a deeper understanding of the underlying mechanisms (Casado-Aranda et al., 2018; Motoki et al., 2020; Yoon et al., 2012). Second, instead of measuring consumer emotions as an outcome of the decision-making process, we measure emotions that consumers experience in the course of the decision-making process. In other words, although self-report measures of consumer emotions can assess the effect on consumers' post decision-making, consumers' emotive states while they undertake the decision-making process can only be assessed through non-intrusive psychophysiological tools (such as automatic detection of facial emotions) used in the current research. Further, our findings also differ from those of Tang et al. (2017) as follows: While the authors found consumers' need of cognition to be an important moderator for positive as well as negative emotions, the current research shows that larger sets not only lead to an amplification of both – negative as well as positive emotions – but these negative and positive emotions also result from two different underlying processes. Specifically, we show that an individual's need for cognition (NFC) impacts their positive affect. On the other hand, the negative affect experienced by consumers during decision-making is driven by the choice difficulty that they experience as a consequence of larger item-sets.

Overall, the current research contributes to the literature on choice overload in three important ways. First, it contributes to the debate on "the existence of choice overload" using objective psychophysiological measures. Using an automated facial expression software, we classify consumers' emotions when they select a product from a large assortment (vs. a small assortment), into discrete emotions to assess if larger assortments lead to adverse effects among consumers. Given the contradictory findings of past research on the existence of choice overload (Chernev et al., 2015; Scheibehenne et al., 2009, 2010), the current research investigates this phenomenon using an alternate method of theory testing – that of assessing participants' real-time psychophysiological reactions.

Second, reconciling past findings on choice overload, our results suggest that consumers experience increased positive as well as negative emotions as they choose from a larger choice set (vs. a smaller set). Specifically, we show that consumers experience an increase in the discrete emotion of “happiness” as well an increase in the negative emotion of “anger” when they choose from a larger assortment (vs. a smaller assortment). This novel finding aligns with the previous research from Iyengar and Lepper (2000), who noted that consumers enjoy selecting from a larger assortment, but also exhibit negative outcomes such as frustration and choice deferral while selecting from these larger assortments. In other words, our findings suggest that “choice overload” does not only lead to an adverse consumer affect, but rather a concoction of positive and negative emotional reactions. This finding may also help explain the reason why Scheibehenne et al. (2010) found a mean effect size of zero, when they averaged the positive and negative outcomes of the effects of product selection from a large assortment.

Finally, answering calls from researchers to delve deeper into the underlying process of choice overload (Chernev et al., 2010), our research notes the existence of two different and parallel mechanisms that explain the simultaneous positive and negative emotions that consumer experience while selecting from a large assortment set. We show that choice difficulty mediates the effect of assortment size on anger, leading to increased anger when choice difficulty increases with assortment size. On the other hand, in line with Tang et al. (2017), we also show that consumers with higher NFC experience greater happiness while selecting from a larger choice set (vs. a smaller set).

1.2 Theory and Hypotheses Building

"Paradox of Choice"

The idea of "choice overload", that is, the notion that an increase in the number of options available to consumers can result in unintended adverse outcomes, has perplexed researchers for decades. Challenging traditional wisdom, in their seminal article, Iyengar and Lepper (2000) showed that while a tasting booth of 24 jams attracted more consumers than a booth with 6 jams, the booth with the larger selection of jams witnessed much fewer actual purchases. Literature on this topic has further explored this phenomenon through the lenses of various moderators. For instance, Chernev (2003) showed that larger assortments may benefit consumers who have had a chance to articulate their preferences, presenting them with a better opportunity of mapping a product with their "ideal" preferences. On the other hand, Diehl and Poynor (2010) note that, as compared to smaller assortments, larger assortments may result in enhanced feelings of dissatisfaction. Specifically, building upon the expectation-disconfirmation theory (Oliver, 1993), the authors show that larger sets may result in excessive expectations of consumers' finding an ideal match, resulting in enhanced feelings of dissatisfaction when these expectations are not fulfilled (Diehl & Poynor, 2010). Overall, literature on the "paradox of choice" is replete of such paradoxes, wherein research has shown that larger sets can result in positive outcomes in some conditions and negative outcomes in others.

Consumer Emotions and Choice Overload

The literature shows that consumers' affective states can influence consumption behaviors in several ways. For instance, consumers' affective states have been noted to impact purchase behaviors (Sherman et al., 1997; Tang et al., 2017), product choice (Garg et al., 2007)), and even cognitive resources available for processing information for decision-making (Mano, 1999). For instance, Tang, Hsieh, and Chiu (2017) show that positive emotions lead to increased purchase intentions, whereas negative emotions lead to reduced purchase intentions. On the other hand, differentiating between the negative emotions of boredom and distress, Mano (1999) showed that both these negative emotions can drive purchase behaviors, but through varying underlying processes. Specifically, while both these negative states drive purchases and alleviate negative emotions, boredom leads to an increase in cognitive capacity, resulting in deliberation of decisions, while distress reduces deliberation (Mano, 1999). Thus, it is critical to understand consumers' discrete emotions in the purchase stage, since different emotional states require divergent managerial interventions.

While past research on choice overload has investigated several important variables such as satisfaction with choice (Chernev, 2003, 2006; Diehl & Poynor, 2010; Iyengar & Lepper, 2000) and choice deferral (Iyengar & Lepper, 2000; Townsend & Kahn, 2014), scant attention has been paid to emotions in this context. Regret, as a consequence of the decision-making is one of the few consumer emotions that has been studied (Haynes, 2009; Inbar et al., 2011). For instance, Inbar et al. (2011) note that for time-constrained decisions, consumers experience feelings of regret after selecting from a larger assortment set.

In one of the rare studies on choice overload that looks at consumer emotions during decision-making, Iyengar and Lepper (2000) noted that the existence of a positive as well as negative consumer affect. Specifically, participants in this study reported experiencing enhanced enjoyment while selecting from a large set, as well as increased perception of negative consequences such as frustration and choice difficulty. In other words, the authors show that the existence of positive emotions does not imply the absence of negative emotions and vice-versa.

Similar thoughts about the existence of mixed emotions – that is the simultaneous occurrence of positive as well as negative emotions – in consumption experiences have been echoed by Ramanathan and Williams (2007). Using the context of indulgent consumption, the authors argue that if consumers experienced only negative emotions such as regret and guilt after engaging in indulgent consumption, they would not exhibit the same behaviors time and again (Ramanathan & Williams, 2007). Thus, considering that past research has often noted the existence of negative outcomes such as regret, frustration, and dissatisfaction as a consequence of excessive choices (Diehl & Poynor, 2010; Haynes, 2009; Inbar et al., 2011), we argue that consumers as well as retailers would have preferred only smaller assortments, if these emotions were the only outcomes of choosing from a larger assortment, indicating the existence of some form of positive emotion that is indeed associated with larger assortment sets.

We argue that the variety-seeking nature of consumers not only attracts them to larger assortment sets, but also makes the decision-making process more enjoyable (Iyengar & Lepper, 2000; Kahn & Wansink, 2004; Tang et al., 2017). In other words, although larger sets can lead to varied negative outcomes, they can also lead to positive emotions and outcomes such as enjoyment (Iyengar & Lepper, 2000) and an improved opportunity of finding an "ideal product" (Chernev, 2003). Thus, we suggest that a larger assortment will lead to an increase in the positive as well as negative emotions that consumers experience during this decision-making process. Formally,

H1: When selecting from a larger assortment set (vs. a smaller assortment set), consumers will experience an increase in the magnitude of positive as well as negative emotions experienced in the decision-making process.

Assortment Size and Choice Difficulty

Prior research has often noted an increase in the level of cognitive costs associated with a larger assortment set (Chernev, 2003; Iyengar & Lepper, 2000; Reutskaja et al., 2018). For instance, in their research using an fMRI and eye tracking, Reutskaja et al. (2018) show that brain activity, as well as saccade frequency increase as the assortment size increases, indicating an increase in the cognitive costs associated with the mental processing of the larger assortment. Similarly, in his research, Chernev (2003) concluded that while selecting from larger sets, consumers with less developed attribute preferences experienced increased decision difficulty as they face a two-step cognitive task : 1) Processing the attributes of a large assortment set and 2) Selecting a product from this assortment. In other words, an increase in the number of product alternatives leads to an increase in the cognitive resources that need to be expended to select a product from such an assortment.

Thus, we suggest that the negative outcomes that result as a consequence of selecting from a larger item-set, such as choice deferral (Iyengar & Lepper, 2000; Shah & Wolford, 2007) and dissatisfaction (Iyengar & Lepper, 2000; Chernev, 2003) are often a consequence of the increased choice difficulty that results as a function of the cognitive resources that larger assortments require for mental processing. In other words, given that human cognitive resources are limited (Baddeley, 1994; Miller, 1956), larger assortments can lead to increased cognitive demands, leading to enhanced perceptions of choice difficulty. Similar findings have been noted in the context of "information overload", wherein research has noted the negative consequences of presenting consumers with excessive information in the form of product alternatives and/or attribute information (Jacoby et al., 1974; Malhotra, 1982). Building on the above findings, we posit that choice difficulty, resulting from an increase in cognitive demands required for processing larger choice sets, will lead to negative emotions in consumers. However, given that choice difficulty is possibly not enjoyable for any human being, we do not expect choice difficulty to explain the variance in positive emotions. Instead, we hypothesize a different mechanism for positive emotions, which we discuss in the following section.

H2a: The increase in negative emotions as a result of a larger choice set (vs. a smaller choice set) will be mediated through consumers' perceived choice difficulty.

Need for Cognition (NFC)

Individual characteristics have been noted to result in varied outcomes in the context of choice overload (Chernev, Böckenholt, & Goodman, 2015). Previous research has investigated several important variables in this context such as decision goals of an individual (Chernev, 2006; Reutskaja et al., 2018), familiarity with the product category (Mogilner et al., 2008), as well as an individual's personal characteristics (Chowdhury et al., 2009; Tang et al., 2017). For instance, Reutskaja et al. (2018) note that when selecting from a large assortment set, participants differed in terms of their brain activity when they pursued a browsing goal as compared to a product-selection goal. On the other hand, Chowdhury et al. (2009) show that maximizers – individuals who always strive to make the best possible decision (Schwartz et al., 2002) – are not only more likely to explore more options when choosing from a larger set, but are also more likely to switch their choices if given an opportunity to do so. In other words, individual factors impact the effect of choice set, such that some consumers may benefit from them while others may not. In the current study we look at one such individual factor – need for cognition (NFC).

NFC, defined as an individual's tendency to engage in and enjoy cognitive endeavors (Cacioppo et al., 1984, p. 48), has been studied in the context of choice overload by a few researchers. For instance, Tang et al. (2017) note that when consumers with a high NFC select from larger sets, they experience increased positive emotions. Similar findings were noted by Chien-Huang and Wu (2006) who show that consumers with higher NFC prefer larger item-sets and are less likely to switch from their original selections when choosing from a larger assortment set (vs. a smaller set).

We suggest that NFC is related to positive emotions, but not to negative emotions due to the fundamental difference between those with higher NFC and those with lower NFC. Difference in NFC does not necessarily mean that consumers with lower NFC expend less cognitive effort than their counterparts with higher NFC (Cacioppo & Petty, 1982). For instance, in a study by Cacioppo and Petty (1982), both groups – those with higher as well as those with lower NFC – reported similar levels of mental effort required for complex problems. However, we argue that although both these groups – consumers with high and low NFC – expend similar levels of cognitive effort while selecting from a larger set, those with a higher NFC prefer larger assortments due to the motivation that drives them to pursue exertive cognitive tasks. Given that individuals with higher NFC are intrinsically motivated to derive positive outcomes (Cacioppo & Petty 1982), although choosing an option from a larger choice set may indeed be cognitively draining for such individuals as well, their motivation to attain positive outcomes will make larger sets more enjoyable for them.

Intuitively, one of the important benefits of a larger consumption set is increased variety and the positive outcomes that may result from this variety (Kahn & Wansink, 2004; Mogilner et al., 2008). For instance, Kahn and Wansink

(2004) show that an increase in variety leads to an increase in anticipated consumption utility. In other words, larger variety can lead to perceptions of enhanced anticipated positive outcomes. We suggest that it is this anticipation of positive outcomes that differentiates consumers with higher NFC from others, and ultimately makes them enjoy decision making from larger assortments. Thus, we expect that NFC will moderate the effect of assortment size on positive emotions, such that consumers with a higher NFC will experience increased positive emotions when selecting from larger assortments.

H2b: The increase in positive emotions as a result of a larger choice set (vs. a smaller choice set) will be moderated by consumers' NFC, such that consumers with higher NFC (vs. lower NFC) will experience more positive emotions.

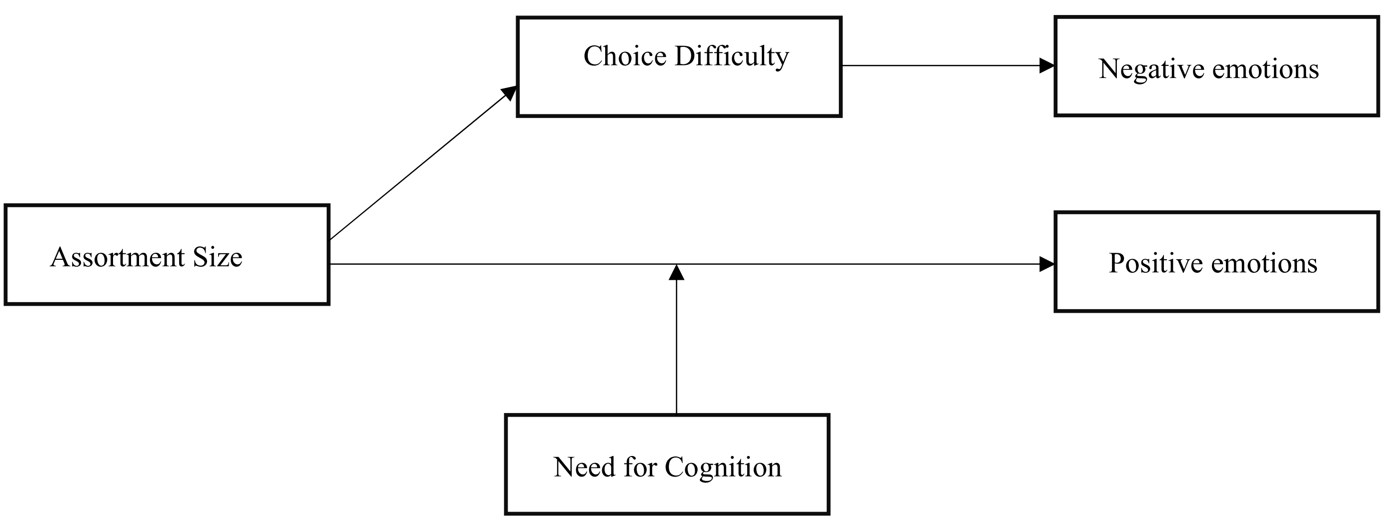


Figure 2. Conceptual Framework of Essay 1

1.3 The Study

Design & Procedure

Forty-nine (49) participants (65% females, M_{age} = 23) participated in a single factor within-subjects design with two levels of choice sets (6 products vs. 24 products). Participants were tasked with selecting a suitable laptop for themselves. In both conditions, participants were given a total of 3 minutes for selecting a product, which was deemed as sufficient based on previous literature (Reutskaja et al., 2018). Following procedures from previous studies which note that less defined preferences can lead to choice overload (Chernev, 2003; Diehl & Poynor, 2010), participants were required to note down their importance weights for the eight attributes of the laptops. Next, participants were shown either a set of 6 vs. 24 laptops, arranged in a matrix format (columns represented eight different attributes and rows represented the various products available), randomized in order. To reduce any bias, all the products were fictitious and were identified using numerical product IDs. After each product selection task, the participant was asked to complete a short self-report questionnaire which served as a manipulation check, and provided a brief break between the two conditions (Frank et al. 2019).

Measures and Apparatus

Discrete facial emotions were collected using Noldus' Facereader (Version 6). FaceReader analyzes 30 frames/second, meaning 30 observations were collected every second. For the current analysis, measures were averaged for each decision period of the task, as is common in psychophysiological studies (Atalay et al., 2012; Reutskaja et al., 2018). FaceReader is able to detect various discrete emotions such as happiness, sadness, anger, surprise, disgust, and neutral. All these emotions are measured on a scale of 0 to 1, wherein 0 indicates that the emotion is absent and 1 indicates that the emotion is fully present. Given that happiness is the only discrete positive emotion that FaceReader analyzes, we use happiness as the positive emotion in the current study. Similarly, considering that anger is the closest negative emotion to frustration (Rydell et al., 2008)– which the prior research has studied (Iyengar & Lepper, 2000) – anger was also analyzed in the current study.

Following past literature, the time taken by the participant to make the decision was used as a proxy for choice difficulty (Chandler & Sweller, 1991; Goodman et al., 2013). The time utilized by participants for each decision was automatically recorded by the stimuli website and was measured in seconds. After completing the study, participants were asked to complete an 18-item scale for NFC (Cacioppo et al., 1984). Unless otherwise stated, a seven-point Likert scale was used for all measures.

Results

We examined the participants' responses to the question concerning whether the number of choices available to them was very limited or very large. A paired t-test indicated that participants in the 24 item-set perceived the assortment to be significantly larger as compared to the 6 item-set (M_{24} = 6.47, SD = .87 vs. M_6 = 3.39, SD = 1.32; t (48) = 15.44, p = .000), indicating that our manipulation was successful.

For testing the impact of assortment size on anger, a one-way repeated measures ANOVA with anger as the dependent variable and assortment size as the independent variable was carried out. In line with H1, the results indicated that participants exhibited more anger while selecting from a larger assortment (vs. a smaller assortment) set, although this result was only marginally significant (M_{24} = .044, SD = .06 vs. M_6 = .032, SD = .04; $F(1, 48) = 3.35$, $p = 0.07$). Further, a repeated measures ANOVA with happiness as the dependent variable showed that participants experienced significantly more happiness in the larger set as compared to the smaller assortment set (M_{24} = .088, SD = .12 vs. M_6 = .069, SD = .08; $F(1, 48) = 5.27$, $p = 0.03$). In line with H1, our results thus indicated that participants experienced an increase in anger as well as happiness simultaneously while selecting from a large assortment.

Figure 3. The Effect of Assortment Size on Consumers' Emotions

To test H2a, we examined if the effect of assortment size on anger is mediated through a resulting increase in the choice difficulty. MEMORE (Model 1, with 10,000 bootstrap samples) was used to test the within-subject mediation effect

(Montoya & Hayes, 2017), wherein anger and choice-difficulty, both were repeated measures. The results indicated that assortment size had a significant effect on choice difficulty (effect = 37.39, $p = .00$, 90% CI [28.25; 46.56]). Further, choice difficulty had a significant effect on the observed anger (effect = .0005, $p = .01$, 90% CI [.0002; .0008]), indicating that an increase in choice difficulty led to increased anger among participants. The indirect effect of choice difficulty on anger was significant after controlling for the effect of assortment size (effect = .018, 90% CI [.006; .03]). However, after controlling for the indirect effect of assortment size on anger through choice difficulty, the direct effect became insignificant (effect = -.006, $p > .50$, 90% CI [- .02; .009]), indicating that choice difficulty fully mediated the effect of assortment size on anger, supporting H2a.

To test if participants' NFC moderated the effect of assortment size on happiness, MEMORE (Model 2, with 10,000 bootstrap samples) was used (Montoya & Hayes, 2017). MEMORE (Montoya & Hayes, 2017) extends the methods for probing an interaction effect between a repeated measures factor (happiness in this study case) and a between-subjects moderator (NFC in this case), that were outlined by Judd et al. (2001). For the purpose of this analysis, NFC was mean centered. The results indicated that NFC had a significant effect on happiness ($\beta = .02$, $t = 2.02$, $p = 0.05$, 95% CI [.00; .05]). We used Johnson-Neyman technique, to identify the range of NFC for which the effect of assortment size on happiness was significant. This analysis revealed that there was a significant positive effect of assortment size on happiness for participants with NFC higher than 4.81 ($\beta = .02$, $t = 2.01$, 95% CI [.00; .03]), but not for participants with NFC lower than 4.81.

The results thus indicate that NFC moderates the effect of assortment size on happiness, supporting H2b.

1.4 Discussion and Conclusion

The results show that larger choice sets tend to amplify positive as well as negative emotions simultaneously. Specifically, using an automatic facial emotion detection, we show that while selecting from larger choice sets, consumers experienced increased happiness, as well as anger at the same time. Further, we uncover two different underlying processes for these varying emotions. In line with previous literature, we show that while positive emotions are a function of an individual's NFC (Chien-Huang & Wu, 2006; Tang et al., 2017), choice difficulty leads to enhanced negative emotions in consumers while they select from larger sets.

Theoretical Contributions

Contributing to the debate on the existence of choice overload (Chernev et al., 2015; Scheibehenne et al., 2010), we use psychophysiological measures to assess the emotions that consumers experience while selecting from large assortments. We note that larger sets lead to amplified mixed emotions, wherein consumers experience more positive as well as more negative emotions at the same time. This finding also partially explains Scheibehenne et al.'s (2010) results wherein they found no effect of assortment size on consumer responses, when they performed a meta-analysis of 50 previous studies.

We also shed some light on the two different phenomena, which are responsible for two divergent outcomes – that of happiness versus anger experienced while selecting from larger sets. For instance, while Iyengar and Lepper (2000) have noted that larger choice sets result in positive outcomes such as increased enjoyment coupled with undesirable outcomes such as choice deferral and frustration, the process underlying this phenomenon remains obscure. Researchers have called for an improved understanding of this phenomenon (Chernev et al., 2010), given the paradoxical findings in this context. Though the current research is limited in several ways, it contributes to a better understanding of this juxtaposition. Specifically, we conclude that choice difficulty drives the outcome of anger, which is often observed among consumers who are faced with sizable assortments. While past research in the context of choice overload has noted that positive emotions lead to purchase intentions (Tang et al., 2017), the emotional cause of negative behaviors (such as choice deferral) that are often associated with choice overload remains underexplored. Our findings suggest that larger sets result in increased anger due to an escalation in choice difficulty, which could possibly explain the negative behaviors that may result with an increased choice.

Managerial Implications

The current research provides several implications for practice. First, instead of focusing on only negative or positive outcomes due to larger assortments, the current research takes a holistic view of this phenomenon to elucidate that larger assortment are double-edged swords – which result in benefits as well as drawbacks. Our findings suggest that larger assortments lead to experiences of positive emotions among consumers with high NFC. Hence, by implementing strategies such as slashing their assortment size (Jeffries, 2015), retailers may deprive such consumers of the happiness that they experience as a result of variety.

On the other hand, our findings suggest that retailers need to adopt more creative ways to reduce choice difficulty, as a larger set also draws increased ire from consumers. For instance, retailers could implement presentation formats that automatically lead to lower perceptions of choice complexity. As an example, Townsend and Kahn (2014) show that verbal depiction of information leads to lower perceptions of choice difficulty. Thus, our findings suggest that retailers need to adopt strategies that give perceptions of increased variety and larger assortments, but do not increase choice difficulty at the same time. For instance, retailers could capitalize on implicit recommendation agents, to decrease consumers' choice difficulty, yet maintain a vast variety.

Finally, given that consumers with higher NFC experience increased positive emotions with a larger item set, retailers, especially in an online context, can use cues to identify such consumers and present them with larger assortments than others. For instance, while a retailer may present all consumers with a default view with a limited assortment, time spent on product information details may serve as a cue to consumers' higher NFC (given that such consumers deliberate on information (Cacioppo et al., 1984), and additional products could be recommended to these consumers in order to expand their assortment set.

Limitations and Further Research

This study limits the assortment sizes to two – a set of 6 versus a set of 24. Although these two sizes of choice sets have been extensively used in the literature (see Chernev et al., 2015 for details), future research could extend them to assess to assortments of multiple sizes. The current research is also limited in terms of the product category used. For instance, it may be possible that consumers experience varied emotions for larger assortments of hedonic products, as compared to utilitarian products such as those used in the current study. Since past research has often noted the differences in consumer behaviors that arise as a result of product type (Senecal and Nantel 2004), the effect of product type on choice overload remains understudied.

Finally, future research on choice overload could examine varied behavioral outcomes that are driven by the two emotions studied in the current research. For instance, are consumers more likely to defer their choice as their anger surpasses their happiness and vice-versa? The optimal balance of these emotions with respect to the assortment set could be a fruitful avenue for future research.

Conclusion

Using psychophysiological measures, the current research contributes to the debate on the existence of choice overload. Overall, our research shows that consumers experience an amplification of both positive and negative emotions as a result of an increase in the assortment size. Further, by elucidating the factors that lead to this intensification of emotions, this work will guide retailers in optimizing the size of their product assortments.

References

Atalay, A. S., Bodur, H. O., & Rasolofoarison, D. (2012). Shining in the center: Central gaze cascade effect on product choice. *Journal of Consumer Research*, *39*(4), 848–866.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*(2), 353–356.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307.

Casado-Aranda, L.-A., Liébana-Cabanillas, F., & Sánchez-Fernández, J. (2018). A neuropsychological study on how consumers process risky and secure E-payments. *Journal of Interactive Marketing*, *43*, 151–164.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332.

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, *30*(2), 170–183.

Chernev, A. (2005). Feature complementarity and assortment in choice. *Journal of Consumer Research*, *31*(4), 748–759.

Chernev, A. (2006). Decision focus and consumer choice among assortments. *Journal of*

Consumer Research, *33*(1), 50–59.

Chernev, A., Böckenholt, U., & Goodman, J. (2010). Commentary on Scheibehenne, Greifeneder, and Todd choice overload: Is there anything to it? *Journal of Consumer Research*, *37*(3), 426–428.

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, *25*(2), 333–358.

Chien-Huang, L., & Wu, P. (2006). The effect of variety on consumer preferences: The role of need for cognition and recommended alternatives. *Social Behavior and Personality*, *34*(7), 865–875.

Chowdhury, T. G., Ratneshwar, S., & Mohanty, P. (2009). The time-harried shopper: Exploring the differences between maximizers and satisficers. *Marketing Letters*, *20*(2), 155–167.

Diehl, K., & Poynor, C. (2010). Great expectations?! Assortment size, expectations, and satisfaction. *Journal of Marketing Research*, *47*(2), 312–322.

Frank, D. W., Stevens, E. M., & Versace, F. (2019). A neurophysiological measure of reward sensitivity and its association with anhedonia in psychiatrically healthy adolescents and young adults. International Journal of Psychophysiology, 141, 56–64.

Garg, N., Wansink, B., & Inman, J. J. (2007). The influence of incidental affect on consumers’ food intake. *Journal of Marketing*, *71*(1), 194–206.

Goodman, J. K., Broniarczyk, S. M., Griffin, J. G., & McAlister, L. (2013). Help or hinder? When recommendation signage expands consideration sets and heightens decision difficulty. *Journal of Consumer Psychology*, *23*(2), 165–174. https://doi.org/10.1016/j.jcps.2012.06.003

Haynes, G. A. (2009). Testing the boundaries of the choice overload phenomenon: The effect of number of options and time pressure on decision difficulty and satisfaction. *Psychology & Marketing*, *26*(3), 204–212.

Inbar, Y., Botti, S., & Hanko, K. (2011). Decision speed and choice regret: When haste feels like waste. *Journal of Experimental Social Psychology*, *47*(3), 533–540.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*(1), 63–69.

Jeffries, Stuart (2015), “Why Too Much Choice Is Stressing Us Out,” The Guardian, , October 21, https://www.theguardian.com/lifeandstyle/2015/oct/21/choice-stressing-us-out-dating-partners-monopolies.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*(2), 115.

Kahn, B. E., & Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *Journal of Consumer Research*, *30*(4), 519–

533.

Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of Consumer Research*, *8*(4), 419–430.

Mano, H. (1999). The influence of pre-existing negative affect on store purchase intentions. *Journal of Retailing*, *75*(2), 149–172.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Mogilner, C., Rudnick, T., & Iyengar, S. S. (2008). The mere categorization effect: How the presence of categories increases choosers' perceptions of assortment variety and outcome satisfaction. *Journal of Consumer Research*, *35*(2), 202–215.

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6–27.

Motoki, K., Suzuki, S., Kawashima, R., & Sugiura, M. (2020). A combination of self-reported data and social-related neural measures forecasts viral marketing success on social media. *Journal of Interactive Marketing*, *52*, 99–117.

Oliver, R. L. (1993). Cognitive, affective, and attribute bases of the satisfaction response. Journal of consumer research, 20(3), 418-430.

Ramanathan, S., & Williams, P. (2007). Immediate and delayed emotional consequences of indulgence: The moderating influence of personality type on mixed emotions. *Journal of Consumer Research*, *34*(2), 212–223.

Reutskaja, E., Lindner, A., Nagel, R., Andersen, R. A., & Camerer, C. F. (2018). Choice overload reduces neural signatures of choice set value in dorsal striatum and anterior cingulate cortex. *Nature Human Behaviour*, *2*(12), 925–935.

Rydell, R. J., Mackie, D. M., Maitner, A. T., Claypool, H. M., Ryan, M. J., & Smith, E. R. (2008). Arousal, processing, and risk taking: Consequences of intergroup anger. *Personality and Social Psychology Bulletin*, *34*(8), 1141–1152.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2009). What moderates the too-much-choice effect? *Psychology & Marketing*, *26*(3), 229–253.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*(3), 409–425.

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, *83*(5), 1178–1197.

Shah, A. M., & Wolford, G. (2007). Buying behavior as a function of parametric variation of number of choices. *PSYCHOLOGICAL SCIENCE-CAMBRIDGE-*, *18*(5), 369.

Sherman, E., Mathur, A., & Smith, R. B. (1997). Store environment and consumer purchase behavior: mediating role of consumer emotions. *Psychology & Marketing*, *14*(4), 361–378.

Tang, Y.-C., Hsieh, Y.-C., & Chiu, H.-C. (2017). Purchase decision: does too much choice leave us unhappy? *European Journal of Marketing*, 1248–1265.

Townsend, C., & Kahn, B. E. (2014). The "visual preference heuristic": The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *Journal of Consumer Research*, *40*(5), 993–1015.

Yoon, C., Gonzalez, R., Bechara, A., Berns, G. S., Dagher, A. A., Dubé, L., Huettel, S. A., Kable, J. W., Liberzon, I., & Plassmann, H. (2012). Decision neuroscience and consumer decision making. *Marketing Letters*, *23*(2), 473–485.

Chapter 2
Dynamic Recommendations: Recommendations when you Need them the Most and by the Right Sources

Abstract

Given the ever-increasing product choices available in the marketplace, usage of recommendation systems is more prevalent than ever. While several online retailers use adaptive, context-aware recommendations that are presented to consumers in the midst of decision-making, these recommendations could theoretically lead to two differential outcomes. One stream of research dictates that such recommendations would ease consumer’s decision-making task by reducing the consideration set. On the other hand, a second stream of research indicates that such adaptive recommendations that are presented in the midst of consumer decision-making could instead further enhance choice difficulty by widening the consideration set. Using neurophysiological tools, we develop and test the implications of a novel, personalized adaptive recommendation system. The findings indicate that recommendations that are presented in the midst of decision-making process alleviate cognitive load. The current research also investigates framing implications for such adaptive recommendation systems.

2.1 Introduction

Given the excessive choices that today's marketplace offers, recommendation systems have become more useful and commonplace than ever. Each year, on an average, 30,000 new products are introduced in the consumer-packaged goods industry itself (Behrmann 2019). It is no surprise that consumers are increasingly relying upon recommendation systems – defined as, "software agents that elicit the interests or preferences of individual consumers for products, either explicitly or implicitly, and make recommendations accordingly" (Xiao & Benbasat 2007, p. 137) – to navigate such plethora of choices. For example, in 2013, a study by McKinsey estimated that 35% of Amazon's purchases were driven by its recommendation system, and this number was as high as 75% for Netflix (MacKenzie, Meyer, & Noble 2013). With recent advances in artificial intelligence (AI) and machine learning, businesses are also adopting technologically advanced recommendation systems that personalize recommendations for each consumer. Netflix, for instance, uses an AI driven recommendation system that takes into account each consumer's viewing habits, preferences, etc. to recommend content that is tailored to the individual.

We note that while recent marketing literature has contributed to many aspects of recommendation systems such as the effect of granularity between products that are recommended (Tsekouras et al., 2020), framing of recommendations (Gai & Klesse, 2019), the effect of product type on recommendations (Longoni and Cian 2020), but the research on personalization in recommendation systems has been few and far between (Chung et al., 2016;

Kawaguchi et al., 2019). The current research contributes to the stream of the literature on personalized recommendation systems to explore how consumers react to recommendations that are personalized based on an individual's choice difficulty.

Literature has shown that whether consumers experience difficulty in selecting a product from a huge assortment largely depends upon several individualistic factors such as the individual's need for cognition, product expertise, and decision strategy being applied (see Chernev et al. 2015 for a meta-analysis). Thus, the personalized recommendation system that is implemented in the current study adapts itself and presents recommendations (from within the initial assortment), based on an individual's choice difficulty, as assessed by their real-time cognitive load. Recommender systems like these are a subset of adaptive personalization systems, since they adapt themselves over time, based on consumers' inputs (Chung et al., 2016). Adaptative personalization systems are employed abundantly in practice, such as Amazon's personalized recommender suggesting consumers "we think you may also like this" after a purchase or other online retailers recommending you product alternatives while you are already deciding among a few other options. The current research focuses on the latter case, that is, consumers' reactions to product recommendations by retailers, when such consumers are already engaged in their decision-making process. Research on such personalized recommendations that are suggested during consumers' current purchase session are of importance because while it is commonplace in practice (Lerche et al., 2016), to date, few research in marketing has investigated if such personalized recommendations that occur facilitate or hamper consumers' decision-making process (see Aljukhadar et al., 2012 for an exception)

Building on past research, we note that such recommendations that are presented to consumers when they are in the midst of decision-making process can lead to two contrasting effects: 1) They may interfere with the consumer's (already advanced) decision-making process and further expand the consumer's consideration set, increasing the choice difficulty (Goodman et al., 2013) or 2) They may facilitate the decision-making process by reducing the consideration set and thus the choice difficulty, which is often a result of "too many choices" (Chernev et al., 2015). For instance, consider that Mary, who is looking for a sweatshirt on Amazon uses the website's filtering techniques to generate a set of 25 sweatshirts. However, while she is examining items from this assortment, Amazon presents her with recommendations from within this set, highlighting options that other consumers seemed to prefer. On one hand, given that Mary has already had some time to develop initial preferences, such recommendations may expand her consideration set, and enhance her choice difficulty. On the other hand, these recommendations may serve as a "shortened consideration set", decreasing her choice difficulty. Given the popularity of such recommendations in practice and the extreme divergence of the two possible effects, this question becomes pivotal for research and practice.

The adaptive personalized system used in the current study is built using neurophysiological measures, such that it adapts and presents recommendations to consumers when it assesses these consumers' cognitive load to be high. In other words, it adapts and recommends product options from (within) the choice set, while consumers are in the midst of selecting a product from this choice set. The aim of the current research is to use objective neurophysiological measures to investigate if such recommendations that occur while consumers have already begun their decision-making process ease this or further worsen the difficulty of the task. While such an adaptive system will be difficult to emulate in practice today, the purpose of this research is to employ neurophysiological measures to understand the effect of this practice used commonly by online retailers for suggesting recommendations in the midst of consumer decision-making. For instance, Amazon frequently suggests product alternatives that a consumer may have browsed before, when the website assesses that the consumer is viewing items that belong to this previously browsed product category (Lerche et al., 2016). On the other hand, many smaller retailers as well as news and media companies use session-based recommendations wherein a consumer's clickstream data behaves as an input to the recommendation system and items similar to the currently being viewed are recommended (Hidasi et al., 2015).

Our research is both theoretically as well as practically relevant and makes several contributions to the marketing literature on recommendation systems. First, we contribute to the literature on adaptive personalized systems by implementing and studying a novel recommendation system that adapts to each individual's cognitive load (and thus choice difficulty) in real time and presents recommendations only when the need arises. As noted earlier, while businesses have implemented several of such personalized systems globally, research on this area is scarce.

Second, given that theory dictates two contradictory outcomes – increase in choice difficulty or its reduction – that could result as a consequence of such recommendations, we test these two divergent hypotheses using objective, neurophysiological measures. We note that neurophysiological tools are the only non-intrusive method of assessing consumers' real time cognitive load (and thus choice difficulty) without interfering with the consumer's decision-making process. Using these a system that classifies cognitive load measured by electroencephalogram (EEG) in real-time, results suggest that such dynamic time-recommendations employed by retailers help facilitate consumers' decision-making process by reducing the cognitive load experienced by consumers.

Third, using a follow-up task-based study carried online, we show that while such session-based adaptive recommendations can lead to consumer benefits, the framing of such recommendations is important. Specifically, the personalized adaptive system developed for Study 1 using EEG only presented one type of recommendations – user-based framing that underscores the similarity between consumers (e.g. other consumers similar to you liked…). Using two different types of framings in an adaptive online study, we conclude that user-based framing that highlights recommendations by "other similar consumers" leads to improved consumer outcomes than recommendations that are framed as those personalized for each individual. As noted by Gai and Klesse (2019), altering the framing of recommendations can lead to differential impact on consumers.

Given the significance of framing of recommendations in real-world, our research not only contributes to the scarce literature on framing of recommendations, but also holds importance for practice.

2.2 Literature Review and Hypotheses Building

Recommendation Systems

Recommendation systems are defined as, "software agents that elicit the interests or preferences of individual consumers for products, either explicitly or implicitly, and make recommendations accordingly" (Xiao & Benbasat, 2007, p. 137). As noted by this definition, some recommendation systems explicitly solicit consumers' input for recommending products, whereas others rely on implicit consumer information such as the products which the consumer may have viewed, her previous purchases, or any other consumer information that a system may obtain without explicitly seeking inputs from the consumer.

Broadly, the literature has categorized recommendation systems into two categories. Content based recommendation systems suggest product options based on consumers' preferences for product attributes (Ansari et al., 2000; Campos et al., 2014). A leading example of such a system is Pandora, a subscription-based music streaming business, wherein a team of musicians categorized each piece of music using more than 400 attributes (Deng, 2019). Whenever a consumer plays a song, she's automatically matched with other music that have the similar musical attributes. On the other hand, collaborative filtering techniques suggest products that are preferred by consumers with similar preferences as the target consumer

(Ansari et al., 2000; Campos et al., 2014). Mimicking word-of-mouth recommendations, such recommendations can be suggested only when a few previous consumers' choices have implicitly or explicitly noted by the recommendation system. Amazon and Netflix often rely on this technique for matching consumers with product choices that other consumers with similar tastes may have liked. Recently, hybrid recommenders that combine the techniques are gaining popularity, in order to capitalize on the benefits of both the techniques (Zhang et al., 2019).

Interactional View of Recommendation Systems

As noted in the literature, recommendation systems are inherently dynamic in nature, otherwise they would have no relevance for consumers (Rana & Kumar Jain, 2015). For instance, a recommendation system would be irrelevant for a consumer if it kept on recommending products to a consumer from a product category that she has purchased recently. In other words, recommendation systems are inherently dynamic and context-aware to some extent so as to suggest products that are of value to consumers. However, researchers in machine learning and AI have frequently called for further research on such dynamic properties of recommendation systems (Adomavicius & Tuzhilin, 2011; Lerche et al., 2016; Rana & Kumar Jain, 2015). By enhancing the context-awareness of recommendation systems, these systems can increase their personalization capabilities for each consumer.

In the realm of recommendation systems, context-awareness has been defined as any information that can be used to characterize the situation of the user of the recommendation system (Campos et al., 2014, pp. 71–72). Further, this context may be interactional, such that the contextual features are dynamic and vary as per the consumer's interaction with the system (Adomavicius & Tuzhilin, 2011). For instance, a recommendation system may use the consumer's inputs in real-time (implicit or explicit) to adapt its recommendations for the consumer. A common example of such an interactive, context-aware recommendation system is conversational recommendation systems, in which the recommendation system's output is heavily dependent upon the consumer's explicit input.

In the context of marketing literature, some recent research has studied such context-aware recommendation systems where the context is variable. For instance, Kawaguchi et al. (2019) investigate moderators (time pressure and crowd pressure) for such personalized recommendation systems which are implemented in the vending machines at Japanese train stations. These vending machines take into account several contextual factors such as the time of the day, as well as interactional factors such as the gender and age of the consumers, which are detected in real-time through face recognition systems installed in these machines, to recommend products that best suit the situation (Kawaguchi et al., 2019). In a similar vein, Chung et al. (2016) implemented personalized recommendation systems in the context of mobile news to conclude that such interactive systems that auto adapt to consumers' observed behavior led to increased propensity of the article being read by the consumer. Lastly, a significant amount of literature in marketing has adopted this interactional view of systems in the context of website morphing, where the websites are adapted based on consumers' clickstream data

(Hauser et al., 2009, 2014; Urban et al., 2014). For instance, Urban et al. (2014) showed that such morphing in accordance with consumers' cognitive styles, doubled the clickthrough rate for banners advertisements when these advertisements were placed on relevant web pages for CNET.

Two Rival Explanations

Dynamic, Context-aware Recommendations Reduce Choice Difficulty

The literature has often noted the negative effects of larger choice sets on various outcomes such as choice deferral (Chernev, 2005; Iyengar & Lepper, 2000), and satisfaction (Chernev, 2003; Polman, 2012), regret (Inbar et al., 2011). This phenomenon, that leads to negative outcomes as a consequence of an increase in the size of a choice set, is termed as choice overload (Chernev et al., 2015; Scheibehenne et al., 2010). In other words, while conventional wisdom would predict positive outcomes as a result of an increase in choices available to consumers, the opposite has been observed by past research.

Given that larger choice sets lead to increased difficulty, conventional wisdom would suggest that recommendations that are shown to consumers while they are in the midst of decision-making should narrow down their consideration set. In other words, given that such dynamic, context-aware recommendations have the potential of reducing the consideration set, such recommendations could ease the perceptions of choice overload. For instance, Häubl and Trifts (2000) show that recommendation agents significantly reduce the number of alternatives consumers inspect, thus decreasing such consumers' consideration sets. Given that larger consideration sets not only leave consumers with reduced time for comprehending each product option but also exhaust such consumers' cognitive resources available for such an exhaustive comprehension (Diehl, 2005), recommendation agents that reduce the consideration set could potentially lead to consumer benefits. Similar findings have been noted in the context of information overload, wherein researchers have noted negative consumer outcomes as a consequence of limited cognitive capabilities when consumers are inundated with an increased number of product information or attribute information (Jacoby et al., 1974; Malhotra, 1982).

Overall, this stream of research implies that a smaller consideration set leads to beneficial consumer outcomes. Similar benefits of a reduced consideration set have also been noted by researchers in the context of decision heuristic applied by consumers. For instance, Chernev (2006) noted that when consumers consider selection of assortment and the subsequent product selection from this assortment as dependent upon each other, they are more likely to select smaller assortments, in an attempt to ease their decision. In a similar vein, Besedeš et al. (2015) concluded that sequential tournament architecture – a piecemeal decision-making strategy that lets consumers divide a larger assortment set into smaller ones, enabling them to select a product option from each of these smaller sets, which are them compared among themselves – leads to reduced choice difficulty.

Given this backdrop, this stream of literature suggests that recommendations that are introduced to consumers in the midst of decision making should reduce their consideration set, leading to an ease in decision making. In other words, such recommendations may help consumers reduce the assortment set instinctively, incentivizing them to focus on the recommended product options so as to reduce their effort. As noted earlier, recommendation agents as well as sorting techniques have been noted to reduce consumers' consideration sets, leading to a reduced number of options that are actually inspected by consumers in such cases (Dellaert & Häubl, 2012; Häubl & Trifts, 2000). Terming such search behavior as "searching in choice mode", the Dellaert and Häubl (2012) explain that with such tools, consumers' expected payoff from inspecting more alternatives reduces. That is, the search cost in such cases increases, while the expected payoff decreases, leaving little motivation for consumers to consider product options outside of such recommendations. Building on this stream of research, we suggest that dynamic recommendations that are presented to consumers in the course of their decision-making process should reduce their consideration set, decreasing the cognitive costs associated with such a task.

Past research has often noted the effect of purchase decision making on consumers' affect (Bui et al., 2011; Schwarz, 2000; Tsiros & Mittal, 2000). In the context of choice overload too, significant research has been carried out in this area (Inbar et al., 2011; Iyengar & Lepper, 2000; Tang et al., 2017). For instance, Tang et al. (2017) noted that need for cognition moderates the effect of assortment size on consumers' positive and negative emotions. Similarly, Iyengar and Lepper (2000) noted that consumers experienced increased positive feelings of enjoyment as well as increased negative emotions while selecting from a larger assortment set. In the current research, we thus suggest that considering that dynamic recommendations have a potential to reduce choice difficulty, we expect such recommendations to influence consumers' affect. Specifically, given that choice difficulty should decrease due to the recommendations, we suggest that consumers' affect will be more positive as a result of dynamic recommendations.

Thus, based on the literature that supports the claim for a reduced consideration set that may result due to dynamic recommendations, we suggest:

H1a: Dynamic recommendations presented to consumers in the course of their decision-making process will reduce the cognitive load associated with this task.

H1b: Dynamic recommendations presented to consumers in the course of their decision-making process will improve consumers' affect.

Dynamic, Context-aware Recommendations Increase Choice Difficulty

On the other hand, a second stream of literature suggests that such recommendations may actually increase choice difficulty instead of decreasing them. This line of thinking suggests that since such dynamic recommendations pop-up when consumers have already begun their product evaluation process, they may interfere with consumers' advanced decision-making process. For instance, Chernev (2003) has noted differences in consumer outcomes with respect to consumers' level of development of preferences. Specifically, Chernev (2003) noted that consumers who have a clear idea of their preferences behaved differently as opposed to those who did have an advanced level of preferences. In the current research, one could argue that consumers would have developed a baseline set of preferences and consideration set before they encounter such dynamic recommendations. Thus, such recommendations could potentially expand this baseline consideration set. Recommendations influence consumers' decisions by impacting their consideration sets. Earlier research has noted the benefits of recommendations in static contexts, where recommendations can shorten the consideration set of consumers and lead to beneficial outcomes (Dellaert & Häubl, 2012; Senecal & Nantel, 2004). However, some studies have also noted the disadvantages of recommendations, especially when consumers have had some time to develop an advanced level of preferences (Fitzsimons & Lehmann, 2004; Goodman et al., 2013; Lurie & Wen, 2014). For instance, a study by Goodman et al. (2013) shows that recommendations increased choice difficulty for consumers with more developed level of preferences, by expanding their consideration sets. Specifically, participants Goodman et al.'s (2013) study were presented with products along with recommendation signage for a second set of choices, after they had already had a chance of making an earlier product selection, interfering with their decision-making process since the participants had already had some time to form initial preferences about their choice. In a similar vein, Fitzsimons and Lehmann (2004) showed that participants exhibited resistance to recommendations, once they had had some time to learn about the products before the recommendations were introduced to them.

Thus, this stream of the literature suggests that given such dynamic recommendations occur after consumers have already begun their product selection task, these recommendations can create a conflict between consumers' baseline consideration set and the recommendation options. In other words, given that in such cases consumers have had some time to form more advanced levels of preferences, dynamic recommendations could create a choice conflict, leading to a larger set consideration set that now needs to be processed. Such an enlarged consideration set would require enhanced level of cognitive processing, leading to increased cognitive load.

As noted, consumers experience varied positive as well as negative emotions as result of product selection from larger assortments (Iyengar & Lepper, 2000; Tang et al., 2017). Given that dynamic recommendations can increase a consumer's consideration set, such recommendations have a potential of deteriorating consumers' affective state. In other words, the increased choice difficulty due to such recommendations can lead to a worsening of consumers' affect as a result of these recommendations.

Thus, we suggest:

H2a: Dynamic recommendations presented to consumers in the course of their decision-making process will increase the cognitive load associated with this task.

H2b: Dynamic recommendations presented to consumers in the course of their decision-making process will a) increase negative emotions experienced while making such decisions and b) decrease positive emotions experienced while making such decisions.

2.3 Study 1: A neurophysiological study to test the two contrasting hypotheses

Design and procedure

Twenty-four participants participated in a single factor within-subjects design (cognitive load before recommendation vs. cognitive load after recommendation) wherein their measures for cognitive load were collected in real-time using EEG [1]. Participants were tasked with selecting a suitable laptop for themselves from a set of 24 products, which has been noted as a large assortment size in the context of choice overload (Chernev et al., 2015; Iyengar & Lepper, 2000). Participants were allocated a total of 3 minutes for selecting a product, which was deemed as sufficient based on previous literature on product selection tasks (Reutskaja et al., 2018). Following procedures from previous studies which note that less defined preferences can lead to choice overload (Chernev, 2003; Diehl & Poynor, 2010), participants were required to note down their importance weights for eight attributes of the laptops. Participants were then instructed to answer some basic demographic questions and questions regarding the importance of laptops for them in their daily lives. Given participants were interacting with the recommendation system for the first time, these pre-task assessments were used to induce the perception of trust in recommendations, which is a prerequisite for consumers to adopt recommendations (Benbasat & Wang, 2005; Komiak & Benbasat, 2006). Participants then undertook a baseline task for calibration

[1] Thirty-four participants participated in the current study, but ten participants never reached the threshold for adaptation of the system, and as a result were never presented with recommendations. Hence, these then participants were excluded from the analysis.

purposes of EEG as well as the Threshold Reactive Adaptive Dynamic Spectrum (ThReADS) (Demazure et al., 2019; Karran et al., 2019), which is described below in detail. Next, participants were shown a set of 24 laptops, arranged in a matrix format (columns represented eight different and rows represented of the various products available), randomized in order. In order to reduce any bias, all the products were fictitious identified using product numbers instead of brand names.

Measures and procedure for dynamic recommendations

The EEG data received from each participant were classified into different levels of cognitive load in real-time, using the ThReADS (Demazure et al., 2019; Karran et al., 2019). Based on the initial baseline task used for calibration, ThReADS classified EEG data for each individual into five levels of cognitive load – 0 indicating very low level of cognitive load and 4 indicating very high level of cognitive load. This system generated two classifications of cognitive load per second in real-time, when the participants were engaged in the task. As mentioned, the aim of this system was to present dynamic recommendations, such that they are presented only when participants experience high relative cognitive load compared to their baseline. Thus, using cognitive load as a proxy for choice difficulty, the system labelled 3 products from the assortment as "recommended by other consumers similar to you". This framing of recommendations was chosen since such recommendations that highlight similarity between consumers is commonly used by companies such as Amazon, and has noted to be more effective by recent literature (Gai & Klesse, 2019). Based on a pretest (Freve-Guerin et al., 2020), it was determined that participants experienced high cognitive load when

ThReADS classified cognitive load as level 3 or 4 (i.e. high or very high) at least 14 times in the last 10 seconds (2 classifications/second). That is, the task induced cognitive load was deemed as high when ThReADS classified cognitive load/per second as high or very high 14 times out of the total 20 classifications that were recorded in the last 10 seconds. At this moment, the system adapted the webpage by itself and three products from the assortment set were labeled as "recommended by consumers similar to you". In reality, all participants received product recommendations that were ranked by the system using the multi-attribute decision making technique called Simple Additive Weighting (SAW) (Hwang & Yoon, 1981; Kabassi & Virvou, 2004), based on their inputs to the importance weights for eight attributes of the laptops. Lastly, given that consumers require to process product information to form initial perceptions about a product (Fitzsimons & Lehmann, 2004; Goodman et al., 2013), the system was programmed to not adapt during the first 10 seconds of the task.

For measuring facial emotions, Noldus' FaceReader (Version 6) was used. FaceReader analyzes 30 frames/second, meaning 30 observations were collected every second. For measuring participants' affect with both positive and negative emotions simultaneously, Facereader calculate a measure called valence. Valence deducts the participants' most intense discrete negative emotion (such as anger, sadness, disgust) from the positive emotion of happiness (happiness is the only discrete emotion that FaceReader categorizes). Given that this measure represents positive as well as negative emotions simultaneously, it can range from -1 to +1, with -1 indicating the highest level of negative affect and +1 indicating the highest level of positive affect.

The measures obtained from Facereader and classification of cognitive load were divided into two time periods: Time1 representing the measures obtained before the dynamic recommendation and Time2 representing the measures obtained after the recommendation. For the current analysis, measures of cognitive load and valence were averaged for each decision period of the task, as is the norm in psychophysiological studies (Reutskaja et al., 2018).

Results

Cognitive load. Dynamic recommendations occurred at an average of 24 seconds after participants were introduced to the product selection task, indicating that on an average, participants experienced high or extremely high levels of cognitive load 70% of the time in the past consecutive 10 seconds around this time period. A one-way repeated measures ANOVA with participants' cognitive load at Time1 and Time2 as the dependent variable indicated a significant effect of dynamic recommendation on participants' cognitive load ($F(1, 23) = 4.19$, $p = .05$). Specifically, participants' classifications of cognitive load were significantly higher at Time1 with $M_{time1} = 2.31$. On the other hand, these classifications of cognitive load decreased after recommendations were introduced to participants $M_{time2} = 2.14$. Thus, in line with hypothesis H1a and contrary to H2a, the results indicate that dynamic recommendations that are presented to consumers in the midst of decision-making reduce the cognitive load that is induced by product selection task.

Consumers' emotions. In order to test if such dynamic recommendations resulted in an improved affect, a one-way repeated measures ANOVA with valence as the dependent variable was carried out. The results, although not statistically significant, suggest a positive trend in valence when comparing Time2 with Time1 (M_{time1} = .084 vs. M_{time2} = .094; $F(1, 23) = .1$, p = NS). Thus, we do not find support for the effect of dynamic recommendations on consumers' emotions (neither H1b nor H2b).

Discussion

Thus, Study 1 sheds some light on whether dynamic recommendations that are presented consumers in the course of their decision-making help decision-making or hinder it. Using cognitive load as a proxy for choice difficulty, we show that such recommendations have the potential to reduce consumers' real-time cognitive load. Using two time periods, we observed a decrease in consumers' cognitive load as a result of dynamic recommendations. Further, we notice that consumers' emotions align with these results directionally. Specifically, while consumers' affect seems to be more positive after viewing the recommendations, these results are not statistically significant.

2.4 Study 2: Framing of dynamic, context-aware recommendations

A recent research on recommendation systems has highlighted the impact that recommendation framings can have on consumers (Gai & Klesse, 2019). Specifically, Gai and Klesse (2019) note that user-based framings, such as "Customers who like this also like…" are more effective than item-based framings such as “Because you like this item, you may also like…”. Given that our first study used only one type of framing, that is, the user-based framing, Study 2 focuses on investigating if framing of dynamic recommendations can result to varied consumer outcomes in terms of consumer affect.

Recommendations based on personalization versus those based on similarity of consumers

Past research in recommendation systems has often noted the impact of personalization on consumers’ behaviors (Dabholkar & Sheng, 2012; Häubl & Trifts, 2000; Kramer et al., 2007; Senecal & Nantel, 2004). For instance, Senecal an Nantel (2004) showed that recommendations personalized by the recommendation agent fared better than recommendations suggested by other consumers or experts. In a similar vein, Häubl and Trifts (2000) noted the benefits of recommendation agents when such agents recommended personalized product suggestions for consumers. Finally, Dabholkar and Sheng (2012) conclude that consumers’ participation in such personalized recommendations drive the consumer benefits, which past literature have noted.

However, some recent research has now noted the impact of recommendations that are based on other consumers’ inputs (Chung et al., 2016; Gai & Klesse, 2019). Specifically, Gai and Klesse (2019) conclude that user-based framing of recommendations – those product options that are preferred by consumers similar in taste with the focal consumer – are more effective since these offer a sort of double-guarantee when consumers believe that their tastes match with “these other consumers”. In a similar vein, Chung et al. (2016) demonstrates that incorporating one's peers' preferred news articles in an adaptive recommendation system improved the system's performance due to social influence and perceptions of similarity.

In the current research, we argue that dynamic recommendations that highlight the similarity between the focal consumer and other consumers will lead to a greater positive affect among consumers. First, we note that such recommendations are presented only when consumers have spent a considerable amount of effort, and time on the purchase decision, without having a final decision outcome. The lack of decision in these cases depict the deficiency of expertise of such consumers with the product in question. As noted by Gai and Klesse (2019), recommendations framed in terms of other consumers are especially effective when consumers lack consumption experience with the product category. Given that such dynamic recommendations are presented to overcome consumers' indecisiveness with unfamiliar products, recommendations from other consumers who have similar preferences and tastes will be more effective and lead to a greater positive affect, since they will offer consumers a sort of warranty against the uncertainty. Second, given that these dynamic recommendations are presented to consumers at a point in decision-making when they are indecisive, social influence can enhance their decision-making capabilities. For instance, Patalano and Wengrovitz (2007) show that indecisive individuals are more confident about their decisions, when working in a group. Thus, we suggest that recommendations from other similar consumers will lead to more positive affect for consumers, since their indecisiveness will be overcome by recommendations from similar consumers. Thus, we suggest:

H3: Consumers' positive affect as result of decision-making will be higher for consumers' when they are presented with dynamic recommendations that are framed in terms of other similar consumers (vs. personalized recommendations by the system).

Study 2

Design and procedure. Sixty-five participants were recruited on an online platform to participate in single factor between-subjects experiment (recommendation framing: other similar consumers vs. personalized recommendation). The procedure for this study was similar to Study 1 expect the following three changes. First, given the inability of measuring real-time cognitive load for adaptation, all participants were shown product recommendations after 24 seconds, since the average time for adaptation in Study 1 was 24 seconds. Second, all participants were shown the same product recommendations, however the framing of the recommendations varied as per the condition to which the participants were randomly allocated to. Lastly, participants' positive affect was measured using a scale adapted from Tang et al. (2017), after they had completed the product selection task.

*Results***.** A one-way ANOVA indicated that the effect of recommendation framing on positive affect was significant ($F(1, 64) = 4.37$, $p = .04$). Specifically, participants who were presented with recommendations framed in terms of other similar consumers led to increased positive affect as compared to those labelled as personalized by the recommendation system ($M_{similar\text{-}consumers}$ = 5.60 vs. $M_{personalized}$ = 5.02). Thus, in line with H3, the results indicate that dynamic recommendations framed as those recommended by similar consumers lead to an increased positive affect as compared personalized recommendations.

2.5 General Discussion and Conclusion

Although dynamic recommendations that are presented to consumers during their decision making are commonplace in practice, few prior research has investigated the effect of such recommendations. Given the two contrasting outcomes that such recommendations can lead to – an increase or decrease in choice difficulty, results of Study 1 suggest that such recommendations have the potential to lead to positive consumer outcomes. Using objective measures obtained through a neurophysiological tool, we show that such recommendations decrease consumers' choice difficulty. Further, although not statistically significant, the measures obtained with automatic facial expression analysis point into the direction of improved affect due to such recommendations. Further, Study 2 shows that the importance of framing in such adaptive recommendations. Specifically, recommendations from other consumers similar to the focal consumer led to a greater positive affect as compared to personalized recommendations. While previous research on static recommendations has noted the opposite result (Senecal & Nantel, 2004), we discuss the contradictory findings in detail below.

Theoretical Contributions

Our research makes several theoretical contributions. First, our research contributes to the limited literature on adaptive personalized systems (Chung et al., 2016; Hauser et al., 2009, 2014). We extend this literature by implementing a novel system that relies on consumers' real-time cognitive load to adapt and suggest recommendations. The EEG-based ThReADS system used in the current study are capable of measuring consumers' neurophysiological responses non-intrusively (Demazure et al., 2019), and adapting the product options for better consumer outcomes.

Second, the current research uses neurophysiological measures to assess an important question: Do such adaptive recommendations hinder or facilitate decision making? This question is pivotal for research and practice, given the two opposing outcomes that such recommendations can lead to. While one stream of research has highlighted the benefits of recommendations (Dellaert & Häubl, 2012; Häubl & Trifts, 2000), other research has noted the detrimental effects of recommendations (Fitzsimons & Lehmann, 2004; Goodman et al., 2013). The current research notes that such recommendations have the potential to reduce choice difficulty among consumers.

Third, our research explores the effect of recommendation framing in the context of dynamic recommendation systems. We note that the current findings, which show that recommendations from similar consumers lead to better consumer outcomes, are contradictory to previous research (Senecal & Nantel, 2004). However, there exist several important differences between the two studies.

First, the recommendations in current research were automatically presented to consumers when they reached a certain threshold of task difficulty, as opposed to Senecal and Nantel (2004), wherein consumers could choose to view recommendations at any point in time. Second, in line with recent literature (Chung et al., 2016; Gai & Klesse, 2019) that has noted the importance of similarity between consumers with respect to recommendations, the current research framed recommendations in terms of “similar other consumers”. Given that consumers today are increasing more accustomed to such recommendations by “similar other consumers”, the current results could have been driven by such acclimatization of consumers (to this framing) over the years.

Managerial Implications

The current research leads to many implications for practice. First, we adapted and employed an innovative personalized system that adapts to consumers’ perceived choice overload and presents recommendations only when consumers need them. Recent technological advances have made pupil and facial recognition systems a part of consumers’ daily lives. For instance, while Apple’s iPhones integrate facial recognition software, Lenovo’s recent laptops have integrated eye-tracking technology in them (Hachman, 2019). Given these technological abilities of everyday devices, retailers may be able to integrate the ThReADS system (Demazure et al., 2019; Karran et al., 2019)to adapt online retail environments for improved consumer outcomes. Further, ThReADS can be adapted to work with consumers’ clickstream data, which is very easily accessible to online retailers in real-time. For instance, users’ clickstream data could be used as a proxy for the neurophysiological measures used in the current study, to gauge their cognitive load in real-time, and dynamic recommendations could be presented when these systems assess the cognitive load to be high.

Second, while several online retailers have been implementing context-aware, adaptive recommendation systems in practice, past research in this context has overlooked the implications of these recommendations. The findings of the current research highlight the benefits of such recommendation systems. Specifically, by illustrating that these adaptive recommendations are able to reduce consumers' real-time cognitive load, the current research encourage managers to adopt such adaptive recommendation systems further.

Lastly, our research underscores the importance of framing in personalized, adaptive recommendation systems. In line with recent literature on recommendation systems (Chung et al., 2016; Gai & Klesse, 2019), our research shows that consumers are more sensitive to recommendations from other similar others, as compared to personalized recommendations. Given that recent recommendation systems such as those implemented by Netflix and Amazon employ hybrid recommendation systems that take into account consumers' data as well as that of other consumers' (collaborative filtering), the findings of current research help guide these businesses. Specifically, by showing that recommendations from similar others fare better than personalized recommendations that are based on consumers' own preferences, our findings present a novel and counterintuitive managerial implication.

Limitations and Future Research

The current research is limited in several ways. Past research has identified several individual characteristics such as product expertise, perceived product risk need for cognition, that may influence consumers in the context of recommendation systems (Xiao & Benbasat, 2007). It may be possible that these factors impact the perceived cognitive load of consumers in the case of adaptive recommendation systems. For instance, as noted in Study 1, ten participants never experienced a website adaptation, since these participants did not cross the threshold of cognitive load, that was required for adaptation. It may be worthwhile for future research to investigate the role of individual factors in the context of adaptive recommendation systems.

Further, past research has also noted the significance of situational factors such as the nature of the task at hand (Castelo et al., 2019; Longoni & Cian, 2020) and product characteristics (Senecal & Nantel, 2004), that impact consumers' adoption of recommendations generated by the recommendation systems. Future research could investigate if adaptive systems are more beneficial in certain situations than others. For instance, it might be possible that adaptive systems offer benefits only for utilitarian products, as was the case in the current study.

Lastly, while the current research only investigated two types of framings for recommendations, there are several possible ways of framing recommendations (Fitzsimons & Lehmann, 2004; Gai & Klesse, 2019; Goodman et al., 2013; Senecal & Nantel, 2004). For instance, recommendations by experts have been noted to lead to differential outcomes (Fitzsimons & Lehmann, 2004; Senecal &

Nantel, 2004). Future research could examine if framing recommendations in terms of the system's expertise or in terms of human expertise leads to improved consumer outcomes.

References

Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217-253). Springer, Boston, MA.

Aljukhadar, M., Senecal, S., & Daoust, C. E. (2012). Using recommendation agents to cope with information overload. International Journal of Electronic Commerce, 17(2), 41-70.

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research*, *37*(3), 363–375. https://doi.org/10.1509/jmkr.37.3.363.18779

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 4.

Besedeš, T., Deck, C., Sarangi, S., & Shor, M. (2015). Reducing choice overload without reducing choices. *Review of Economics and Statistics*, *97*(4), 793–802.

Bui, M., Krishen, A. S., & Bates, K. (2011). Modeling regret effects on consumer post-purchase decisions. *European Journal of Marketing*.

Campos, P. G., Díez, F., & Cantador, I. (2014). Time-aware recommender systems: A comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, *24*(1–2), 67–119. https://doi.org/10.1007/s11257-012-9136-x

Chernev, A. (2003). When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, *30*(2), 170–183.

Chernev, A. (2005). Feature complementarity and assortment in choice. Journal of Consumer Research, 31(4), 748-759.

Chernev, A. (2006). Decision focus and consumer choice among assortments. *Journal of Consumer Research*, *33*(1), 50–59.

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, *25*(2), 333–358.

Chung, T. S., Rust, R. T., & Wedel, M. (2009). My mobile music: An adaptive personalization system for digital audio players. *Marketing Science*, *28*(1), 52–68.

Chung, T. S., Wedel, M., & Rust, R. T. (2016). Adaptive personalization using social networks. *Journal of the Academy of Marketing Science*, *44*(1), 66–87.

Deng, H. (2019). Recommender systems in practice. *Medium*. https://towardsdatascience.com/recommender-systems-in-practice-cef9033bb23a

Dabholkar, P. A., & Sheng, X. (2012). Consumer participation in using online recommendation agents: effects on satisfaction, trust, and purchase intentions. *The Service Industries Journal*, *32*(9), 1433–1449.

Dellaert, B. G. C., & Häubl, G. (2012). Searching in choice mode: consumer decision processes in product search with recommendations. *Journal of Marketing Research*, *49*(2), 277–288.

Demazure, T., Karran, A., Labonté-LeMoyne, É., Léger, P.-M., Sénécal, S., Fredette, M., & Babin, G. (2019). Sustained attention in a monitoring task: Towards a neuroadaptative enterprise system interface. In *Information Systems and Neuroscience* (pp. 125–132). Springer.

Diehl, K. (2005). When two rights make a wrong: Searching too much in ordered environments. Journal of Marketing Research, 42(3), 313-322.

Diehl, K., & Poynor, C. (2010). Great expectations?! Assortment size, expectations, and satisfaction. *Journal of Marketing Research*, *47*(2), 312–322.

Fitzsimons, G. J., & Lehmann, D. R. (2004). Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science*, *23*(1), 82–94.

Frève-Guérin, E., Léger, PM, & Sénécal, S. (2020). Development of a neuroadaptive web interface modulated according to cognitive load.

Gai, P. J., & Klesse, A.-K. (2019). Making recommendations more effective through framings: impacts of user-versus item-based framings on recommendation click-throughs. *Journal of Marketing*, *83*(6), 61–75.

Goodman, J. K., Broniarczyk, S. M., Griffin, J. G., & McAlister, L. (2013). Help or hinder? When recommendation signage expands consideration sets and heightens decision difficulty. *Journal of Consumer Psychology*, *23*(2), 165–174. https://doi.org/10.1016/j.jcps.2012.06.003

Hachman. (2019). *Lenovo's yoga S940 intros smart assist eye-tracking, noise filtering, and a glass-wrapped display*. PCWorld. https://www.pcworld.com/article/3331137/lenovo-yoga-s940-intros-smart-assist-eye-tracking-defocusing-features-and-a-glass-wrapped-display.html

Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*, *19*(1), 4–21.

Hauser, J. R., Liberali, G., & Urban, G. L. (2014). Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science*, *60*(6), 1594–1616.

Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). Website morphing. *Marketing Science*, *28*(2), 202–223.

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based recommendations with recurrent neural networks. In Proceedings of the International Conference on Learning Representations. https://arxiv.org/abs/1511.06939

Hwang, C. L., & Yoon, K. (1981). Methods for multiple attribute decision making. In *Multiple attribute decision making* (pp. 58-191). Springer, Berlin, Heidelberg.

Inbar, Y., Botti, S., & Hanko, K. (2011). Decision speed and choice regret: When haste feels like waste. *Journal of Experimental Social Psychology*, *47*(3), 533–540.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*(6), 995–1006.

Jacoby, J., Speller, D. E., & Kohn, C. A. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*(1), 63–69.

Kabassi, K., & Virvou, M. (2004). Personalised adult e-training on computer use based on multiple attribute decision making. *Interacting with computers*, *16*(1), 115-132.

Karran, A. J., Demazure, T., Leger, P.-M., Labonte-LeMoyne, E., Senecal, S., Fredette, M., & Babin, G. (2019). Towards a hybrid passive BCI for the modulation of sustained attention using EEG and fNIRS. *Frontiers in Human Neuroscience*, *13*, 393.

Kawaguchi, K., Uetake, K., & Watanabe, Y. (2019). Effectiveness of product recommendations under time and crowd pressures. *Marketing Science*, *38*(2), 253–273.

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly*, 941–960.

Lerche, L., Jannach, D., & Ludewig, M. (2016). On the Value of Reminders within E-Commerce Recommendations. *Conference on User Modeling Adaptation and Personalization*, 27–35. https://doi.org/10.1145/2930238.2930244

Longoni, C., & Cian, L. (2020). Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The “Word-of-Machine” Effect. *Journal of Marketing*.

Lurie, N. H., & Wen, N. (2014). Simple decision aids and consumer decision making. *Journal of Retailing*, *90*(4), 511–523.

MacKenzie, I., Meyer, C., & Noble, S. (2013, October 1). How retailers can keep up with consumers. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers

Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of Consumer Research*, *8*(4), 419–430.

Patalano, A. L., & Wengrovitz, S. M. (2007). Indecisiveness and response to risk in deciding when to decide. *Journal of Behavioral Decision Making*, *20*(4), 405–424.

Polman, E. (2012). Effects of self–other decision making on regulatory focus and choice overload. *Journal of Personality and Social Psychology*, *102*(5), 980.

Rana, C., & Kumar Jain, S. (2015). A study of the dynamic features of recommender systems. *Artif Intell Rev*, *43*, 141–153. https://doi.org/10.1007/s10462-012-9359-6

Reutskaja, E., Lindner, A., Nagel, R., Andersen, R. A., & Camerer, C. F. (2018). Choice overload reduces neural signatures of choice set value in dorsal striatum and anterior cingulate cortex. *Nature Human Behaviour*, *2*(12), 925–935.

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of consumer research*, *37*(3), 409-425.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*(4), 433–440.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers’ online choices. *Journal of Retailing*, *80*(2), 159–169.

Tang, Y.-C., Hsieh, Y.-C., & Chiu, H.-C. (2017). Purchase decision: does too much choice leave us unhappy? *European Journal of Marketing*, 1248–1265.

Tsekouras, D., Dellaert, B. G., Donkers, B., & Häubl, G. (2020). Product set granularity and consumer response to recommendations. Journal of the Academy of Marketing Science, 48(2), 186-202.

Tsiros, M., & Mittal, V. (2000). Regret: A model of its antecedents and consequences in consumer decision making. *Journal of Consumer Research*, *26*(4), 401–417.

Urban, G. L., Liberali, G., MacDonald, E., Bordley, R., & Hauser, J. R. (2014). Morphing banner advertising. *Marketing Science*, *33*(1), 27–46.

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. MIS quarterly, 137-209.

Zhang, S., Yao, L., Wales, S., Sun, A., Tay, Y. I., & Tay, Y. (2019). Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Comput. Surv*, *52*(5). https://doi.org/10.1145/3285029

Chapter 3
Because AI Can Err but Not Cheat: When and How Anthropomorphism of AI Technology Alters Consumers' Responses

Abstract

Recent advances in AI and machine-learning have led to an increased adoption of personified, anthropomorphic AI-enabled technological agents such as Apple's Siri and Amazon's Alexa. Given the pervasiveness of such anthropomorphic AI-enabled technological agents in consumers' daily lives, the current research investigates consumers' reactions to instances when these AI agents fail. Specifically, distinguishing between the types of failures as performance-based and benevolence-based failures, the results show that consumer responses to such anthropomorphic agents depend upon the type of failure. While consumers exhibit ameliorated responses to performance failures attributed to anthropomorphic (vs. non-anthropomorphic) agents, the same does not hold true for benevolence failures. Benevolence failures attributed to anthropomorphic (vs. non-anthropomorphic) AI-agents elicit exacerbated consumer responses. Further, the results show that these differences in consumer responses arise due to consumers' varying perceptions about the cause of the failure. While performance failures attributed to anthropomorphic (vs. non-anthropomorphic) AI-agents are seen as more "accidental", benevolence failures attributed to such agents are perceived as more "intentional" violations.

3.1 Introduction

With recent advances in artificial intelligence (AI), AI powered systems are being used pervasively in every field of life. From self-driving cars and conversational chatbots that answer consumers' mundane queries to digital voice assistants such as Amazon's Echo speakers or Google's Nest speakers, AI and machine learning technologies are increasingly helping consumers in various everyday tasks of life, often replacing the need of human diligence. Such is the iniquitousness of AI, that in a recent study by Accenture (2019), 50% of global online consumers indicated that they already use some form of digital voice assistants, whereas another 14% affirmed that they planned on purchasing such an agent within the next year.

With increased consumer interactions with such AI backed systems, retailers are increasingly adopting personification elements such as the endowment of a name, gender, personality, or humanlike physical attributes to anthropomorphize such digital agents (Borau et al., 2021; Purington et al., 2017). Given that research has shown that such anthropomorphism of AI technology leads to several benefits such as increased acceptance and engagement (Borau et al., 2021; Li et al., 2010), it is no surprise that Amazon' virtual assistant technology named Alexa (employed by Echo speakers) and Apple's Siri (employed in by Apple's phones, tablets, etc.), which are anthropomorphized by virtue of gender and name, are now part of millions of consumers' lives.

With the growing widespread usage of AI-enabled systems, failures involving such agents are inevitable. While recent marketing research on AI has often investigated various factors that facilitate or inhibit consumers' adoption of such technology (Castelo et al., 2019; Dietvorst et al., 2018; Gai & Klesse, 2019; Leung et al., 2018; Longoni & Cian, 2020), research on consumers' reactions when such systems fail is in its infancy (Dietvorst et al., 2015; Gill, 2020; Srinivasan & Abi, 2021).

The current research examines when and why personification of AI systems salvages or exacerbates consumers' negative reactions to failures attributed to these agents. This question is of pivotal importance for managers who are increasingly personifying AI-driven agents in an attempt to harness the benefits of increased acceptance and engagement that results from such personification (Borau et al., 2021; Li et al., 2010).

Using four studies, we provide evidence that consumers use varied heuristics when faced with failures attributed to personified and non-personified AI agents. While personified AI-agents are subject to normative standards of conduct that are generally used to judge transgressions attributed to humans, such norms are not applied to non-personified agents. In other words, we suggest that consumers automatically use internalized belief systems of moral norms that are commonly used to assess human failures, to adjudicate failures attributed to "humanlike" personified AI systems, but not to non-personified systems. Specifically, distinguishing between the type of failures into performance-based failures (defined as failures that result as a lack of competency or skills) and benevolence-based failures (defined as failures when a system fails to care about and act in the interests of the consumer), we show that compared to non-personified systems, people are more tolerant to performance-based failures committed by personified AI systems. By contrast, in case of benevolence-based failures, consumers exhibit aggravated negative reactions when such failures are attributed to personified AI systems.

Our research is both theoretically novel and substantively impactful. Our first set of theoretical contributions relates to research on failures that are attributed to AI-backed technology (Dietvorst et al., 2015; Gill, 2020; Srinivasan & Abi, 2021). While consumers' failed experiences have been studied extensively in traditional contexts such as hospitality and travel, researchers have increasingly called for research that investigates the role of technological agents in the context of failures (Khamitov et al., 2020). We extend the research on failures attributed to AI-enabled technological agents by exploring how and when personification of such agents (Kim et al., 2019; Mende et al., 2019; Purington et al., 2017), leads to ameliorated or exacerbated consumer reactions.

Table 1. Summaries of Relevant Research on Technological Failures and Anthropomorphism

Research	Method	Main Independent Variable(s)	Moderator(s)	Main Findings
Fan et al. (2016)	Scenario-based experiment	Service failure with an anthropomorphic vs. robotic voiced airline self-service technology	• Presence of other customers • Consumers' sense of power	• Consumers' perceptions of power, and the presence of other consumers in the situation interact with anthropomorphism of the technology, to lead to differential consumer outcomes.
Choi et al. (2020)	Scenario-based experiments	Service failure in varied contexts (restaurant/hotel) by a humanoid vs. non-humanoid robot	Failure type: process vs. outcome	• Increased dissatisfaction with humanoid (vs. non-humanoid) robots for process failures. • No difference was observed in satisfaction with humanoid (vs. non-humanoid) robots for outcome failures. • The effect for process failures is mediated by lower warmth perceptions.
Fan et al. (2020)	Scenario-based experiment	Service failure with an anthropomorphic vs. machinelike	• Consumers' self-construal (measured)	• Anthropomorphism leads to decreased dissatisfaction with the failure.

Research	Method	Main Independent Variable(s)	Moderator(s)	Main Findings
		airline self-service technology	• Consumers' self-efficacy (measured)	• Anthropomorphism interacts with interdependent self-construal and self-efficacy to lead to differential outcomes.
Lin et al. (2020)	Online experiment	Conflict of recommendations from virtual salesperson with high vs. low automated social presence of avatar	• Cuteness of virtual agent	• Negative effects of conflict are mitigated when the virtual agent's avatar has high automated social presence • A virtual agent with a more (vs. less) cute avatar is not able to ease the conflict.
Srinivasan and Sarial-Abi (2021)	Online experiment	Brand harm crises caused by an anthropomorphic vs. non-anthropomorphic algorithm	• No moderator	• More negative responses when the error is caused by an anthropomorphized algorithm

Second, by introducing an important moderator in this context – the type of failure defined by consumers' perceptions of the cause of the failure – we help reconcile the mixed findings of how anthropomorphism of technological agents leads to better or worse consumer outcomes, when such agents fail. ***Table 1*** summarizes the primary results of empirical studies that have examined the effect of anthropomorphism in the context of technological failures/conflicts. Some research in this context (Choi et al., 2020; Srinivasan & Abi, 2021) have shown that anthropomorphism leads to worsened negative responses from consumers, others have shown that anthropomorphism of technological agents leads to improved responses in the context of failures and conflicts (Fan et al., 2020; Lin et al., 2020). Categorizing failures based on consumers' perceptions of the cause of failure, we introduce an alternative typology of technological failures and reconcile the above findings to show that consumers exhibit increased negative responses such as anger and a desire to avoid using such technological agents when these agents commit benevolence-based failures. However, this finding does not hold true for performance failures, wherein anthropomorphized AI agents fare better than their non- anthropomorphized versions.

Third, building on deontic justice, the 'justice rules' that guide moral treatments of others (Cropanzano et al., 2017), we dissect the underlying process for these differential outcomes of failures by AI-enabled technological agents. Specifically, we show that performance-based failures by personified agents trigger greater perceptions of "accidental betrayals" – defined as "regrettable errors by the trustee, such that the trustee had no intentions of violating the consumer's expectations" (Elangovan & Shapiro, 1998, p. 551) – as compared to intentional betrayals (defined as "betrayals that consumers view as intentional violations of their expectations by the trustee" (Elangovan & Shapiro, 1998, p. 551). On the other hand, benevolence-based failures committed by such personified agents elicit greater perceptions of "intentional betrayals". This effect driven by the differential betrayals demonstrate that consumers are particularly hold personified AI agents to higher moral standards. In other words, given that consumers hold personified agents to humane standards of morals, while performance failures are overlooked as "accidents", self-interested failures by these agents are viewed more negatively.

3.2 Theoretical Development

Anthropomorphism

Anthropomorphism, defined as "seeing the human in non-human forms and events" (Aggarwal & McGill, 2007), has been studied extensively by recent research on branding (Aggarwal & McGill, 2007, 2012; Mourey et al., 2017; Puzakova & Aggarwal, 2018; Puzakova & Kwak, 2017). Literature in this context has often noted how attributing humanlike features, such as name, gender, or physical characteristics, to brands and objects can lead to varied consumer behaviors towards such brands and objects. For instance, Mourey et al. (2017) show that engaging with anthropomorphized products can satisfy consumers' need for social belonging. On the other hand, underscoring the importance of moderators in this context, Puzakova and Kwak (2017) demonstrate that while consumers prefer interactive anthropomorphized brands in socially uncrowded situations, they show a reduced preference for such anthropomorphized brands in socially crowded situations. Overall, literature in this context has concluded that anthropomorphizing brands and products results in consumers behaving toward such brands and humans in a manner that they would with humans. From reacting to such anthropomorphized partner (or servant) brands as they would with human partners (or servants) (Aggarwal & McGill, 2012), to withdrawing from such anthropomorphized brands in a socially crowded situation (as they would with other humans) (Puzakova & Kwak, 2017), anthropomorphization results in consumers overlooking the reality that these brands and objects are indeed non-human. Below we discuss how similar findings have been noted in research on anthropomorphism in the context of technology.

Anthropomorphism and Technology

Given the "automated" nature of AI-enabled agents and the requirement of consumers' social interaction with such agents for these agents to function, they are inherently anthropomorphic in nature to some extent (Purington et al., 2017; Van Doorn et al., 2017). For instance, while personifying Amazon's Echo with the name "Alexa" further induces anthropomorphism (Lopatovska & Williams, 2018; Purington et al., 2017), the automated nature of Echo and the interaction between a consumer and the agent themselves create a primitive level of anthropomorphism. Terming this phenomenon as "automated social presence", Van Doorn et al. (2017, p. 44) note that such AI-enabled technological agents "make consumers feel that they are in the company of another social entity". To further enhance these agents' social presence, businesses are increasingly infusing anthropomorphism in technology through personification elements such as name, gender and personality (Borau et al., 2021; Purington et al., 2017; Van Doorn et al., 2017). Such are the advancements in personified AI applications, that students at Georgia Tech university were surprised to know that their teaching assistant, Jill Watson, was an AI-enabled agent and not a real human (Kaplan & Haenlein, 2019).

Overall, past research on anthropomorphism in AI-enabled agents has shown that such anthropomorphism leads to beneficial consumer outcomes (Kim et al., 2019; Li et al., 2010; Lopatovska & Williams, 2018; Purington et al., 2017). For instance, Purington et al. (2017) noted that personification of AI-agents leads to more sociable interactions and increased consumer satisfaction. In a similar vein, a qualitative study by Lopatovska and Williams (2018) noted that consumers often exhibit mindless politeness toward such personified, anthropomorphic agents, by using words like "thank you" and "please" in their interactions with such agents.

Considering that past research on anthropomorphism in branding and technology suggests that consumers often treat anthropomorphic entities as humans, below we discuss how consumers are predisposed to employ humane moral principles of deontic justice to judge transgressions committed by these humanlike anthropomorphic agents.

Deontic Justice and Failures Attributed to AI

Research on deontic justice indicates that humans are guided by moral norms of social conduct and feel obliged to uphold them (Cropanzano et al. 2003; Folger et al. 2005). This sense of morality, which views that justice is important for its own sake, leads people to evaluate a transgression with respect to some 'normative criteria' or 'justice rules' (Cropanzano et al. 2017). These justice rules, which are guided by moral norms of human interaction, rather than by the consequence of the transgression itself, dictate if the violation is perceived to be unjust or no. For instance, an unfair outcome could be the result of an intentional moral violation or could be an unintentional miscalculation. While the former case would lead to moral outrage, the latter would not violate any "justice rule". These principles of moral norms are deeply ingrained in people, and as in other forms of heuristics, humans often exhibit automatic judgements and responses to moral transgressions of these norms (Cropanzano et al. 2003).

Given this backdrop of deontic justice, we suggest that in case of failures attributed to AI-enabled agents, anthropomorphism of the agent will dictate how rigorously consumers uphold these moral norms of justice, to judge transgressions committed by these agents. As noted earlier, attributing humanlike characteristics to non-human entities intuitively leads consumers to behave with such entities like they would with 'actual humans' (Aggarwal & McGill, 2012; Lopatovska & Williams, 2018; Puzakova & Kwak, 2017). In other words, given the humanlike characteristics of these anthropomorphic entities, consumers unconsciously treat such anthropomorphic agents as humanlike. In this regard of attributing deeper human characteristics to nonhuman agents, Waytz, Heafner, and Epley (2014) note that anthropomorphizing a non-human agent does not only involve attribution of superficial characteristics to it, but rather entails attributing it with essential characteristics of humans, such as mind and ability to think. Thus, given that anthropomorphizing an AI-agent instinctively leads consumers to believe that these anthropomorphic agents possess humanlike mental capacity to think, consumers are likely to uphold these agents to humane standards of social conduct. That is, given that consumers unwittingly accredit anthropomorphic agents with 'humanlike mental capabilities' (Waytz et al., 2010, 2014), we posit that they automatically judge transgressions of these agents using 'justice rules' of social norms that are generally used appraise human interactions.

Further, given that deontic justice pertains to the morality of a transgression, rather than to the outcome of the transgression per se, below we explain how the type of failure will dictate the application of these moral rules of justice. Specifically, categorizing failure type into two types – performance-based and benevolence-based – below we hypothesize how consumers' reactions to failures attributed to AI-enabled agents will be moderated by this classification of failures.

Performance-based and Benevolence-based Failures

In order to make sense of negative episodes such as transgressions and failures, consumers engage in automatic cognitive and affective appraisals of these situations (Folkes, 1988; Folkman & Lazarus, 1984). Research has noted various primary as well as secondary cognitive dimensions such as attribution of blame, controllability, morality, and ethicality that are used by consumers to judge these negative transgressions (Khamitov et al., 2020). In the current research, we suggest that task-based AI-failures are automatically 'cognitively appraised' by consumers into two categories – performance-based failures and benevolence-based failures. Given that accepting and entrusting a technological agent requires placing your trust in its competence as well as benevolence (Benbasat & Wang, 2005; Komiak & Benbasat, 2006; Wang & Benbasat, 2007), we suggest that a failure of the task essentially results into consumers' cognitively reasoning the failure of the task. Based on literature in Information Systems, we define performance-based failures as those wherein the AI-agent fails the consumer due to a lack of ability, skills, or expertise that are essential to perform its task effectively (Benbasat & Wang, 2005). On the other hand, we define benevolence-based failures as those wherein consumers believe that the system failed to care about and act in the interests of the consumer (Benbasat & Wang, 2005).

Research in Information Systems has extensively noted the importance of the two dimensions of trust – performance and benevolence, in adoption of technological agents for task accomplishment (Benbasat & Wang, 2005; Komiak & Benbasat, 2006; Wang & Benbasat, 2007). Using technological agents for tasks such as decision-making, product selection, or autonomous driving requires that consumers trust the agent to not only be capable of accomplishing the task, but also act in the best interests of the consumer. For instance, in a recent research on autonomous cars, Gill (2020) noted that consumers expect these AI-enabled cars to put the drivers' interests over those of the pedestrians, underscoring the importance of both the dimensions of trust – performance and benevolence in task related usage of AI-agents.

Type of Failures, Justice and Anthropomorphism

We suggest that a failure attributed to an AI-enabled agent would lead to cognitive appraisal process such that these failures would be classified into the two defined categories – performance-based or benevolence-based, depending upon consumers' internalized 'justice rules'. While a performance-based failure would mean that a consumer assesses the agent to essentially be incapable of accomplishing the task due to a lack of skill or capability, such failures would not elicit moral outrage. Past research on deontic justice and moral violation has stressed the importance of a transgressor's motivations in this context (Cropanzano et al., 2003; Folger et al., 2005). Specifically, the principles of deontic justice dictate that although performance and benevolence failures may lead to the same outcome, the performance-based failure will not be judged as violating moral norms as these failures are not caused by a transgressor's willful violation of morals.

Further, we suggest that consumers will be more tolerant toward such performance failures attributed to anthropomorphic agents. Given that anthropomorphism of entities leads consumers to attribute deeper, humanlike characteristics such as mental capabilities, personality (Kim et al., 2019; Waytz et al., 2010), we suggest that such anthropomorphic AI-agents will also be attributed with ineptness and lack of competencies, attributes that are generally used to describe humans, not machines. In other words, given that humanlike features of anthropomorphic agents result in consumers associating humanlike beneficial qualities such as trustworthiness to such agents (Waytz et al., 2014), this 'humanness' of such agents will act as a buffer for performance failures attributed to anthropomorphic AI-agents, since consumers inherently consider humans inferior to machines in terms of efficiency and competence (Longoni & Cian, 2020). Secondly, our moral conditioning through societal norms, religion, etc.

makes consumers more accepting of unintentional human errors. For instance, the adage "to err is human" holds true across cultures and generations. Thus, given consumers' inherent conditioning of forgiving unintentional human errors, they are more likely to be easier on anthropomorphic, humanlike AI-agents than on non- anthropomorphic agents.

Thus, we hypothesize:

H1: Consumers will exhibit attenuated negative reactions to performance-based failures attributed to anthropomorphic AI-agents (vs. non- anthropomorphic AI-agents).

Unlike performance failures, AI failures classified as benevolence-based should cause moral outrage among consumers. When a consumer entrusts an AI-agent with a task, there exists an agency relationship between the consumer and the agent, since more often than not, the agent possesses more information than the consumer with respect to the target behavior, leading to a situation of information asymmetry (Wang & Benbasat 2007, p. 221). For instance, when a consumer entrusts an AI-enabled autonomous car with the task of driving, the consumer is generally unaware of the intricacies of machine-learning and AI behaviors that the car has been trained upon. Thus, when a consumer trusts an AI-agent with a task, he/she is implicitly assuming that the AI-agent will place his/her interests above those of other parties (Benbasat & Wang, 2005; Gill, 2020). In other words, by entrusting the AI-agent with a task, the consumer is expecting the AI-agent to uphold the norm of caring about the consumers interests and placing them before those of any other party involved, including the technological agent and its associated parties themselves.

Thus, we suggest that when a consumer appraises a failure to be a benevolence-based one, he/she will assess such a failure to violate the 'justice rule' of social conduct. Further, given that anthropomorphism of entities leads consumers to attribute such agents with humanlike capabilities such as mental ability to think and form intentions (Srinivasan & Abi, 2021; Waytz et al., 2010, 2014), for benevolence failures, this 'mindlike capability' of such anthropomorphic agents will exacerbate consumer reactions. These mindlike capabilities of anthropomorphic agents will lead consumers to believe that given this failure was not a result of lack of capabilities, the agent knowingly and thoughtfully violated the social norms of conduct by failing to act in the consumers' interests. Simply stated, these moral violations of norms will automatically lead consumers to believe that such humanlike anthropomorphic agents, that are capable to think and thus able to uphold social norms, intentionally violated the moral norms.

Past research on transgressions has noted the effect of willful violations with regards to detrimental consumer responses that consumers enact as a reaction to such transgressions (Grégoire et al., 2010; Kähr et al., 2016). For instance, Grégoire et al. (2010) show that when consumers perceive firms' transgressions as an intentional act of greed, they engage in revenge behaviors against such firms, in an attempt to avenge these transgressions. Similar findings were noted by Kähr et al. (2016), who note that consumers who believe they have been wronged by the firm exhibit aggressive hostile behaviors such as sabotaging the firm. Thus, given that benevolence-failures attributed to anthropomorphic AI-agents will be viewed as intentional acts of norm violation, we suggest that consumers will exhibit increased negative reactions against such anthropomorphic agents, whom the consumers attribute with mindlike capabilities.

H2: Consumers will exhibit exacerbated negative reactions as a response to benevolence-based failures attributed to anthropomorphic AI-agents (vs. non-anthropomorphic AI-agents).

Accidental vs. Intentional Betrayal

We suggest that given humans' tendency to treat anthropomorphic agents in the same way as they treat other humans (Waytz et al., 2014), consumers will be more likely to assess performance failures (vs. benevolence failures) attributed to anthropomorphic agents as "accidental betrayals", defined as, "regrettable errors by the trustee, such that the trustee had no intentions of violating the consumer's expectations" (Elangovan & Shapiro, 1998, p. 551). In other words, given consumers' predisposition that humans are prone to accidentally err, we suggest that consumers will be more likely to regard performance failures committed by such humanlike agents as "accidental betrayals". On the other hand, given that consumers hold anthropomorphic agents to higher standards of morals, benevolence-based failures committed by such anthropomorphic (vs. non-anthropomorphic) agents are more likely to be viewed as acts of intentional betrayal (vs. accidental betrayal).

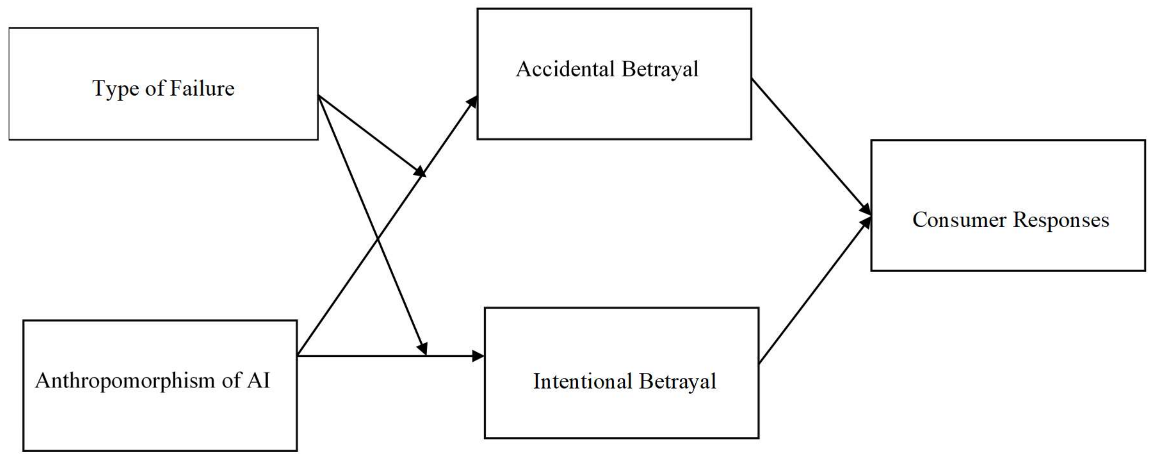


Figure 4. Conceptual Framework of Essay 3

Thus, we expect:

H3a: For performance failures, such failures will be perceived as more accidental (compared to intentional) betrayals when such failures are committed by anthropomorphic (vs. non- anthropomorphic) AI-agents.

H3b: For benevolence failures, such failures will be perceived as more intentional (compared to accidental) betrayals, when such failures are committed by anthropomorphic (vs. non- anthropomorphic) AI-agents.

H4 The effects hypothesized in H1 and H2 will be mediated by these differences in betrayal perceptions. Specifically:

H4a: For performance failures, the amelioration in negative reactions toward anthropomorphic (vs. non- anthropomorphic) AI-agents will be mediated through consumers' perceptions about accidental (vs. intentional) betrayals about these failures.

H4b: For benevolence failures, the exacerbation in negative reactions toward anthropomorphic (vs. non- anthropomorphic) AI-agents will be mediated through consumers' perceptions about intentional (vs. accidental) betrayals about these failures.

Overview of Studies

We tested our hypothesis in a series of four studies in the context of AI-enabled systems. In an empirical field analysis, Study 1 examines consumers' tweets posted on Twitter regarding performance failures that they experienced with AI-enabled virtual assistants. Using a real-service failure in the context of recommendation agents, Study 2 provides experimental evidence that performance failures attributed to anthroponomic agents (vs. non- anthroponomic agents) result in ameliorated negative reactions from consumers, whereas benevolence failures attributed to such agents lead to worsened negative responses from consumers. This study also employs automated facial expression analysis (FaceReader), to analyze consumers' emotions in real time, as a response to the failure. Study 3 uses an alternate dependent variable to replicate the findings of Study 2, and also attempts to understand the underlying process that leads to the differential outcomes outlined earlier. Finally, highlighting the need to differentiate between the two types of failures, Study 4 shows that in the case of extreme consumer reactions to failures – such as consumers' "desire for revenge" to avenge the failure – consumer responses to failures by anthropomorphic (vs. non- anthropomorphic) agents only differ with respect to benevolence failures, but not performance failure.

3.3 Study 1: Analyses of Consumer Tweets

The primary objective of Study 1 is to provide empirical field evidence for our prediction that consumers react to a performance failure less negatively when the failure is attributed to an anthropomorphic AI-enabled agent (vs. a non-anthropomorphic AI-enabled agent). To accomplish this goal, tweets (from the microblogging website Twitter) with hashtags relevant to failures with two of the most popular virtual assistants, that is Google Home and Amazon's Alexa were analyzed . Research has shown that this microblogging website is frequently used by consumers for documenting their positive as well as negative experiences with products and services (Jansen et al., 2009).

Method

Tweets were scrapped by querying the API of Twitter using Python programming language, using hashtags pertaining to failures of Amazon Alexa and Google Home. The tweets (from 1st January 2017 to 27th February 2021) resulted in 523 tweets about failures pertaining to Google Home (Home or Nest) and Amazon's Alexa (Echo). Recent literature has shown that anthropomorphism is induced by personification, such as endowment of a name, gender, and personality to digital agents (Borau et al. 2021; Purington et al. 2017; Lopatovska and Williams 2018). Hence, tweets that attributed failures to Google Home were coded as "non-anthropomorphic", whereas tweets that consisted of failures about Amazon's Alexa were coded as "anthropomorphic". Further, a research assistant was also asked to classify these tweets into performance-based failures and benevolence-based failures. However, given that only 16 tweets were classified as benevolence failures, whereas the vast majority of the tweets were identified as performance-based failures (507 tweets), benevolence failures were not analyzed in the current study.

To operationalize consumers' reactions towards these failures, we analyzed consumers' sentiments using LIWC (Pennebaker et al., 2001). LIWC is a dictionary based automated textual analysis tool and has been routinely employed to measure consumers' underlying affective states (e.g. Dhaouia and Websterb 2021; Berger and Milkman 2012). As recommended by Humphreys and Wang (2018) we used two measures for the analysis of affect in these tweets. Specifically, we measured the emotional tone and negative emotions of the tweets. Emotional tone is a summary variable, and a higher number is associated with a more positive tone (Pennebaker et al. 2015). On the other hand, the variable negative emotions is operationalized in LIWC as the percentage of negative emotional words to the total number of words in that particular tweet. Hence, a higher number on this variable indicates a presence of greater negative emotions.

Results

Pretest. Sixty-three users of Google Assistant or Amazon Alexa were recruited through MTurk to complete a small survey online. Questions were added to ensure that only legitimate users of either one of the devices could participate. Results indicate that Amazon Alexa (vs. Google Homea) users perceive their virtual assistant to be significantly more anthropomorphic ($M_{\text{Amazon-Alexa}} = 4.2$, $SD = 1.81$

vs. $M_{Google=Home} = 3.27$, SD = 1.66, F (1,61) = 4.50, p = .038). No differences were perceived with respect to warmth, competence or brand perceptions (all p's > .10).

Main Results. An ANOVA with tone as the dependent variable and anthropomorphism of the agent as the independent variable indicated a significant effect (F(1, 505) = 67.34, p = .000)[2]. In line with H1, tweets classified as anthropomorphic were more positive in tone ($M_{anthropomorphic} = 53.65$, SD = 39.04 vs. $M_{non\text{-}anthropomorphic} = 26.15$, SD = 35.36). It should be noted that a value above 50 denotes a positive tone, whereas a value below 50 denotes a negative tone (Pennebaker et al. 2015), indicating that performance failures attributed to anthropomorphic agent were positive in tone, whereas those attributed to the non-anthropomorphic agent were negative in tone. Different covariates such as the consumer' followers, as well as the likes, retweets, and responses received to by the tweet were factored in as covariates. However, given that none of the covariates were significant (p > .5 for all covariates), these were eliminated from further analyses.

Similar results were obtained with negative emotions as the dependent variable. Specifically, when the failure was attributed to a non-anthropomorphic AI-agent (vs. an anthropomorphic agent) the tweets contained more negative emotional words ($M_{anthropomorphic} = 1.53$, SD = 2.78 vs. $M_{non\text{-}anthropomorphic} = 5.00$, SD = 4.02, F(1, 505) = 131.32, p = .000). Again, none of the covariates mentioned (the

[2] As noted earlier, only performance-based tweets were analyzed in this study since the majority of the tweets were identified as performance-based (507 tweets were identified as performance-based while 16 were coded as benevolence-based).

consumer' followers, and likes, retweets, and responses received to by tweet) were significant (p > .4), and were eliminated from further analyses.

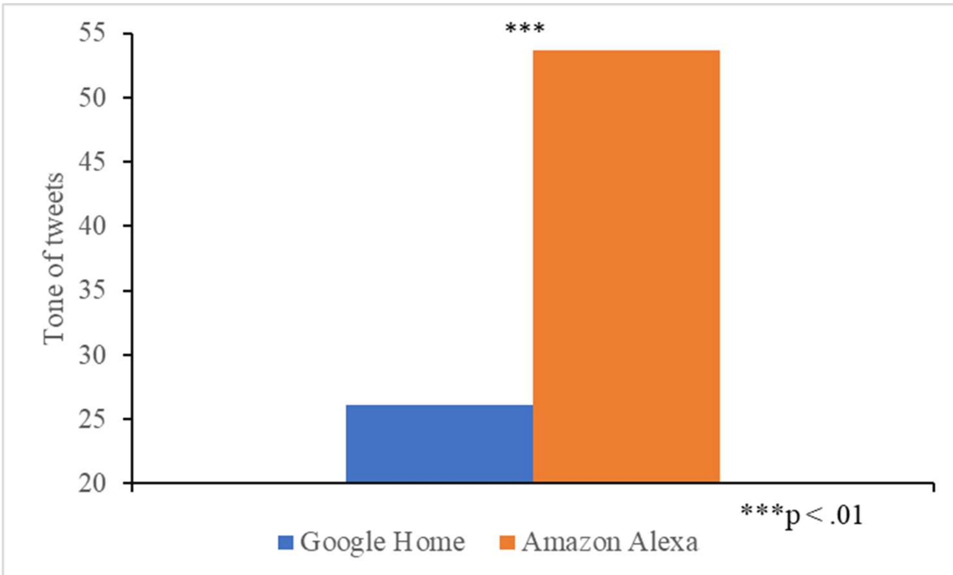


Figure 5. Tone of tweets for Performance Failures of Google Home and Amazon Alexa

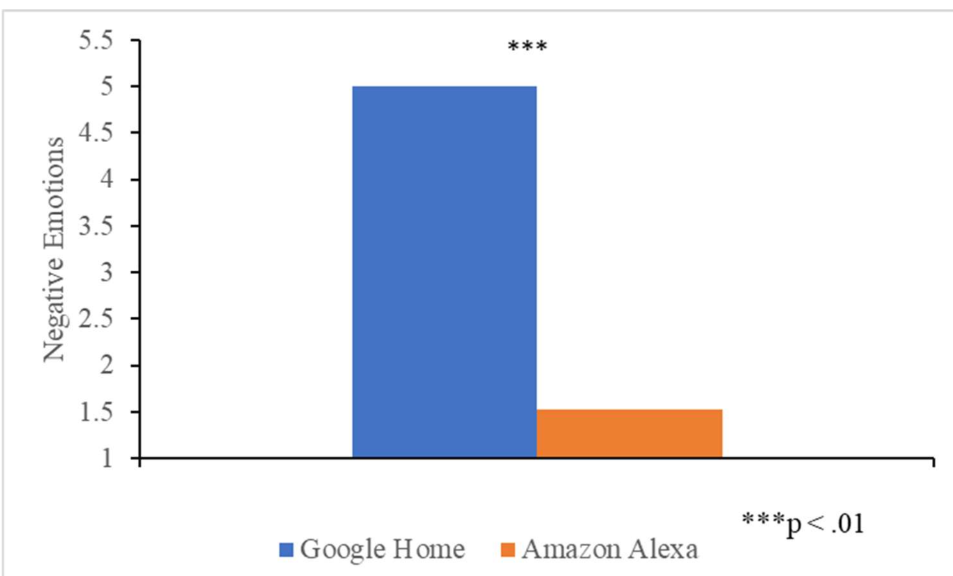


Figure 6. Negative emotional words in tweets for Performance Failures of Google Home and Amazon Alexa

Discussion

Overall, the results of this field study show that for performance-based failures, consumers exhibit decreased negative reactions for anthropomorphic (vs. non-anthropomorphic) agents. These results were replicated using two dependent variables (tone of the tweet and number of negative emotional words), validating the robustness of this finding. An alternative context (recommendation agents), as well as an alternate operationalization for anthropomorphism were also tested using sentiment analysis of consumer tweets, and indicated similar results for the tone of these tweets (see Appendix B for the detailed study). Further, this field study also leads to another important finding: that in the context of AI-agents, performance-oriented failures seem to outnumber benevolence-based failures. Given that research has often noted the benefits of anthropomorphizing AI-agents using various personification agents (e.g., Borau et al. 2021; Purington et al. 2017), the current findings suggest that anthropomorphic AI-agents also have an advantage over non- anthropomorphic agents in a majority of failures, which tend to performance-oriented in nature.

3.4 Study 2: Real Service Failure with Psychophysiological Measures

Study 2 had four main objectives. First, one of the primary objectives of Study 2 was to replicate the findings of Study 1 in a controlled setting. Using a controlled setting in Study 2 eliminates possible confounds of Study 1. Second, Study 2 also extends the previous findings by testing the interaction between anthropomorphization of AI-agents and the type of failure. By manipulating the type of failure, Study 2 tests how consumers' responses to anthropomorphic (vs. non- anthropomorphic) agent vary depending upon the type of failure. A third goal of this study was to validate the hypotheses using a real service failure. Given the criticisms of hypothetical scenario-based studies in the context of service failures (Khamitov et al. 2020), this study aimed to test the external validity of our predictions using a real service failure. Finally, this study employed psychophysiological measures obtained through a FaceReader to assess participants' emotional responses to the service failure. FaceReader, which uses automated facial coding of human expressions, has been shown to be as adept, and in some cases even better than human coders in recognizing human emotions (Skiendziel et al. 2019; Lewinski et al. 2014).

Method and Measures

Method. Participants were recruited through two online platforms (MTurk and Prolific) to participate in a 2 (failure type: performance-based vs. benevolence-based) X 2 (anthropomorphism: yes vs. no) between-subjects design study. A video-recorder was integrated in the study, and participants were walked through essential steps that helped them record themselves as they participated in the study. Given the technical complexity of this study, only participants with at least an undergraduate degree were invited to participate in the study. Further, considering that the study involved a number of different steps, such as working with a chatbot that recommended products, 17 participants were excluded as they faced some form of technical issue or were not able to understand the different steps of the task. The final sample size included of 212 (two-hundred and twelve) participants

(M_{age} = 34; 42.9% females), who were able to successfully complete all the steps of the task, including capturing themselves via the video-recorder.

The stimuli and task were pretested using a series of pretests as outlined below. This study employed a product selection task, wherein participants were asked to select one of the top 5 laptops on the market. Participants were told that they would receive $5 if they were able to select one of the top 5 laptops on the market and only $2.50 if they selected a laptop that was not among these "top 5". The participants were told that to assess the top 5 laptops on the market, the research team would use an objective third-party report, that was available to the team. A seemingly real market research report was designed for this purpose, which listed the "top 5 laptops on the market". It should be noted that this report was not accessible to the participants at the product selection stage, but was introduced to them at a later stage, as outlined below.

Further, given the extensively huge number of laptops available on the market, participants were told that they would receive help from an AI-based recommendation agent, which would work with the participants to help them shortlist these laptops. Specifically, participants were told that the recommendation system would ask them some questions, and then based on these responses, it would give them a list of "top 10 laptops". Participants would then select a laptop from this "top 10 laptops", which would determine if they will receive $5 or $.250 as per the criteria listed earlier.

At this stage, the anthropomorphism of the recommendation agent was manipulated, such that participants in the anthropomorphic condition were introduced to Skylar, an anthropomorphic recommendation agent, personified using humanlike physical attributes. On the other hand, participants in the non-anthropomorphized condition were introduced to chatbot labeled "the recommendation system", which resembled a machine. The stimuli were previously pretested for anthropomorphism with a different sample (see Appendix B for details of stimuli).

All participants then interacted with the recommendations system to enter their preferences on different attributes of a laptop, such as screen size, RAM, etc. They were then presented with 10 laptops, as suggested by the recommendation agent. It should be noted that all participants received the same top 10 recommendations irrespective of their answers to the system. The list of these 10 laptops was pretested, such that these laptops offered significantly less value (in terms of price) as compared to the "top 5 laptops on the market" (see Appendix B for details).

All participants then proceeded to select a laptop. To create a service failure, they were then told that their selected laptop wasn't one of the top 5 laptops on the market since they could have selected a laptop with better value. Thus, the participants would be paid only $2.50. At this stage, they were also introduced to the market report, which showed 5 laptops that were not recommended to the participants earlier, to create an illusion that the system failed them by not recommending any of the "top 5 options" which were significantly lower priced.

Finally, to manipulate the reason of the failure, participants were told that they would be offered some insights into the working of the recommendation agent. For participants in the performance failure condition, they were told that Skylar (vs. recommendation agent) is limited in its capabilities and may have to limit its search if the laptops available on the market are extensive. For participants in the benevolence failure condition, participants were told that the compensation that a brand pays on sales made through Skylar (vs. recommendation agent) is also a factor for determining the recommended options. On the completion of the study, participants were briefed that they would receive the promised $5, irrespective of their choice of laptop. The task was pretested for the type of failure, and to ensure that the participants blamed the system (vs. themselves) for the failure (see Appendix B for details).

*Measures***.** Participants' facial emotions were recorded as they read the reason for the failure, as this time-period represented an interaction of the anthropomorphism of the recommendation chatbot with the type of failure. Participants' videos were analyzed using Noldus FaceReader (version 8), which uses machine learning to automatically analyze discrete facial emotions such as happiness, sadness, anger, surprise, fear and disgust on a scale of 0 (absent) to 1 (fully present). Given that the current study aimed at assessing consumers' negative responses to failure by AI-agents, measures for anger were extracted from the FaceReader. FaceReader has been noted to classify anger with an accuracy of 84% to 96% (Skiendziel et al. 2019; Stöckli et al. 2018).

Further, to triangulate psychophysiological measures with participants' self-reported measures of negative emotions towards the failure, negative affect was measured using a scale adapted from Gregoire et al. (2018) (see Appendix A for measures). A 5-point scale anchored by *not at all* (1) versus *extremely* (5) was used for this measure (Cronbach alpha = 77%). Lastly, unless otherwise mentioned, a seven-point Likert scale was used for all measures used in pretests (see Appendix A for measures and Appendix B for details).

Results

Pretests. One-hundred and nineteen (119) US residents recruited through Prolific completed a pretest for anthropomorphism of the stimuli. The study used a one factor (anthropomorphism: yes vs. no) between-subjects design, wherein participants were asked to observe an image of the recommendation chatbot. A one-way ANOVA with anthropomorphism (measured using one item; see Appendix A for details) as the dependent variable indicated that participants in the anthropomorphic condition perceived the recommendation chatbot to be significantly more humanlike as compared to the non-anthropomorphic chatbot ($M_{anthropomorphic} = 3.69$, $SD = 1.56$ vs. $M_{non\text{-}anthropomorphic} = 3.09$, $SD = 1.60$, $F(1, 117) = 4.34$, $p = .039$). A separate pretest was conducted to compare the value (price) of the "top 10 laptops as recommended" by the recommendation agent and "top 5" laptops on the market. This pretest recruited ninety-one (91) US residents from Mturk for a within-subjects study (laptops: top 10 vs. top 5), wherein participants were allocated to both the conditions in a randomized order. Paired t-tests indicated that the top 5 laptops on the market were rated as significantly higher in terms of value in terms of price ($M_{top5} = 4.76$, $SD = 1.62$ vs. $M_{top10} = 4.38$, $SD = 1.56$, $t(90) = 2.19$, $p = .03$). Finally, the task was pretested using procedures similar to the main study, except that participants were not video-recorded, and were promised a sum of $2 on selecting one of the top 5 laptops (vs. $ 1 for not doing so). Eighty-seven (87) participants recruited via MTurk completed the task without any technical difficulty. The pretest indicated that benevolence (vs. performance) failure was rated significantly higher ($M_{performance} = 3.41$, $SD = 1.62$ vs. $M_{benevolence} = 4.4$, $SD = 1.59$, $F(1, 85) = 8.13$, $p = .005$) on a scale that measured if the system had failed to put their interests above its own (Cronbach alpha = 94.6% ; see Appendix A for items). Finally, a paired t-test with measures for self-blame (Cronbach alpha = 94.6%) and blame on the recommendation system (Cronbach alpha = 94.6%) indicated that the blame on the system was significantly higher ($M_{system\text{-}blame} = 4.41$, $SD = 1.74$ vs. $M_{self\text{-}blame} = 3.48$, $SD = 1.79$, $t(86) = 2.81$, $p = .006$), meaning that the failure was attributed to the recommendation system.

Results of Main Study. Demographic data of participants from both the recruitment platforms (Prolific and MTurk) was analyzed for differences. There were no significant differences for gender and income (both p's > .3) between the participants recruited from the two platforms. However, the participants from Prolific were significantly younger than those from MTurk ($M_{Prolific} = 33.8$ vs. $M_{MTurk} = 38.22$, $F(1, 210) = 7.31$, $p = .007$), and thus age was controlled for in the analyses.

Results with psychophysiological measures. FaceReader was able to generate values for 203 participants for the time-period of interest. Three outliers (standardized scores exceeding $z = +/- 3.29$) were excluded from the analyses of FaceReader data as these measures could be a result of irregularities with this measurement (Harley et al. 2013; Haapalainen et al. 2010), resulting in a total of

200 participants for this measure. A two-way ANCOVA with measures for anger from FaceReader as the dependent variable, and age as a covariate indicated a significant interaction effect of failure type and anthropomorphism ($F (1, 195) = 7.98$, $p = .005$). The effect of age as a covariate was insignificant ($p > .9$), and hence this variable was excluded from further analyses of psychophysiological measures. Supporting H1, planned contrasts indicated that in the case of performance failures, participants' experienced significantly lower anger when such failures were attributed to anthropomorphic agents (vs. non-anthropomorphic agents) ($M_{anthropomorphic} = .099$, $SD = .17$ vs. $M_{non-anthropomorphic} = .18$, $SD = .24$, $F(1, 196) = 4.09$, $p = .045$). On the other hand, for benevolence failures, participants' anger toward the failure was significant marginally, such that they experienced increased anger when such failures were attributed to anthropomorphic agents (vs. non- anthropomorphic agents) ($M_{anthropomorphic} = .16$, $SD = .22$ vs. $M_{non-anthropomorphic} = .08$, $SD = .13$, $F(1, 196) = 3.85$, $p = .05$).

Results with self-report measures. A two-way ANCOVA with self-reported negative affect as the dependent variable and age as the covariate indicated a significant interaction effect of failure type and anthropomorphism ($F (1, 207) = 6.09$, $p = .01$). As with psychophysiological measures, the effect of age was not significant ($p > .9$), and hence was excluded from further analyses of this measure. In line with previous results, planned contrasts indicated that in the case of benevolence failures, participants' experienced more negative affect when such failures were attributed to anthropomorphic agents (vs. non- anthropomorphic agents), although this effect was marginally significant ($M_{anthropomorphic}$ = 1.88, SD = .95 vs. $M_{non\text{-}anthropomorphic}$ = 1.56, SD = .75, $F(1, 208) = 3.56$, $p = .06$). On the other hand, contrasts for performance failures were not significant, although they were directionally in line with the hypotheses. Specifically, participants experienced lower negative affect toward anger when performance failures were attributed to anthropomorphic agents (vs. non- anthropomorphic agents) ($M_{anthropomorphic}$ = 1.63, SD = .88 vs. $M_{non\text{-}anthropomorphic}$ = 1.92, SD = .93, $F(1, 208) = 2.65$, $p = .11$).

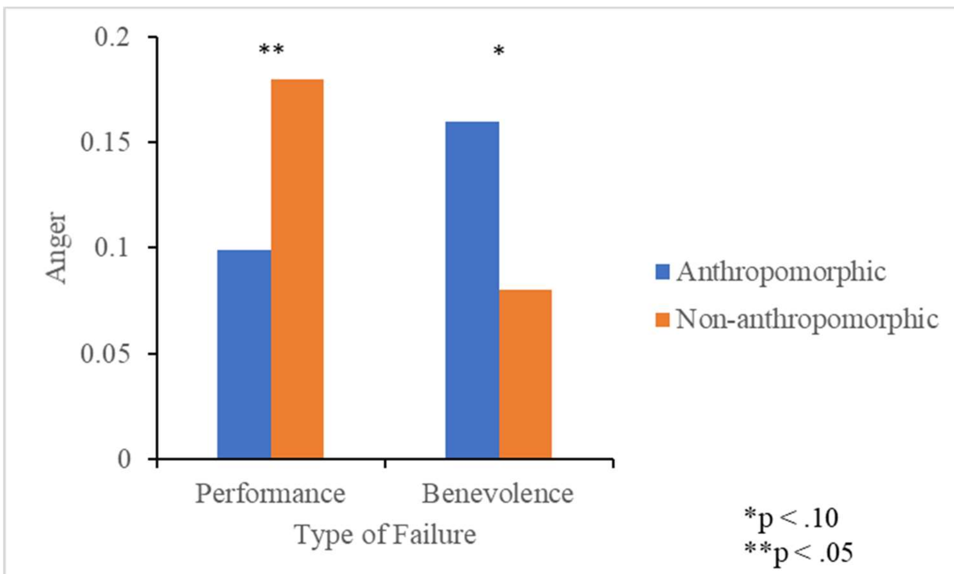


Figure 7. Interaction of Anthropomorphism and Failure Type on Anger measured by FaceReader

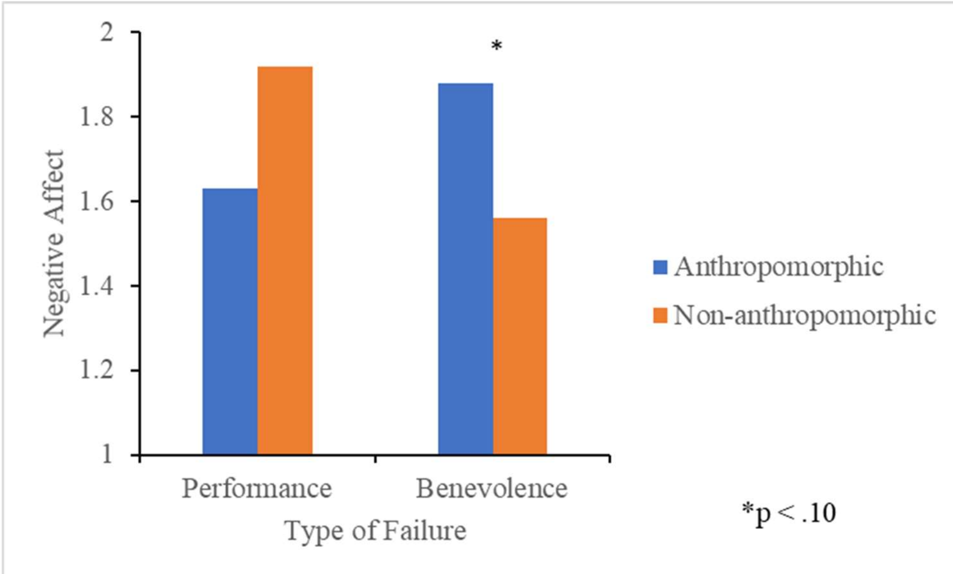


Figure 8. Interaction of Anthropomorphism and Failure Type Self-Reported Negative Affect

Discussion

Overall, using psychophysiological measures obtained via automated coding of consumers' emotional reactions to failures by AI-agents, Study 2 shows that consumers exhibit reduced negative responses to performance failures attributed to anthropomorphic agents (vs. non-anthropomorphic agents). On the other hand, this pattern reverses for benevolence failures, wherein consumers experience increased negative responses to such failures, when they are committed by anthropomorphic (vs. non-anthropomorphic agents). Further, participants' self-report measures were in line with the physiological measures, validating the robustness of the results.

3.5 Study 3: The Underlying Process

Study 3 had two goals. First, this study aimed to replicate the previous findings with a more downstream variable – consumers' desire to avoid the AI agent for future usage. Given that consumers are increasingly entrusting AI agents with tasks around their daily lives, it is likely that consumers will be reluctant to use an AI-agent which fails them. Second, this study aimed at understanding the process that leads to differential consumer outcomes with anthropomorphic agents (vs. non- anthropomorphic agents), depending upon the type of failure.

Method

Four hundred and fifty-three (M_{age} = 36, 43.3% females) recruited via MTurk completed a 2 (failure type: performance vs. benevolence) X 2 (anthropomorphism: yes vs. no) between-subjects design study. Using the same context as Study 2, the current study asked participants to imagine that they were on an online retailer's website, and decided to use the retailer's recommendation system, which asks them some questions about their preferences, and uses algorithms to suggest suitable laptops. Similar to Study 2, at this stage, participants were introduced to an anthropomorphic or non-anthropomorphic version of the recommendation system. Anthropomorphism of the system was manipulated through the appearance of the system, using a previously pretested image of the system that resembled the system used in study 2 (see Appendix B for details). Participants were then told that based on the recommendations of the system, they decided to purchase a suitable laptop. Further, participants were asked to imagine that a week after this purchase, they came across a laptop that suits their needs better but was not recommended by the virtual agent. After deciding to scrutinize the retailer's website, the participants realize that the recommendation agent only recommended limited number of brands. At this stage, the type of failure was manipulated with stimuli similar to Study 2. Specifically, participants in the performance failure were told that the FAQ section of the system states that system is limited in its abilities, whereas those in the benevolence failure were told that the system may give preference to laptops by "preferred partners" (see Appendix B for detailed stimuli).

Measures. Measures were collected for participants' desire to avoid the system using a 3-item (Cronbach alpha = 93%) scale adapted Grégoire et al. (2008) (see Appendix A for items). Lastly, participants' perceptions about intentional betrayal (Cronbach alpha = 89.3%) and accidental betrayal (a = 73.5%) of their trust were also measured. All items were measured using a 7-point Likert scale (1= Strongly disagree and 7 = Strongly agree).

Results

Betrayal Perceptions. For testing H3, a difference score was constructed with measures for participants' perceptions about intentional and accidental betrayal (correlation between intentional and accidental betrayal = – .48, p = .00) such that Δ betrayal perceptions = intentional betrayal – accidental betrayal. Thus, a higher score on this measure indicated that participants' perceptions of intentional betrayal exceeded those of the failure causing an accidental betrayal and vice-versa. As shown in *Figure 9*, a two-factor ANOVA with anthropomorphism and type of failure as independent variables indicated a significant interaction effect

(F (1, 449) = 16.7, p = .000). Specifically, performance failures attributed to anthropomorphic (vs. non-anthropomorphic) agent led to higher perceptions of accidental betrayal as compared to intentional betrayal ($M_{anthropomorphic}$ = -1.51, SD = 2.19 vs. $M_{non-nthropomorphic}$ = -0.39, SD = 2.19, F (1, 449) = 13.12, p = .000). On the other hand, for benevolence failures, such failures attributed to anthropomorphic (vs. non-anthropomorphic) agent led to higer perceptions of intentional betrayal (compared to accidental betrayal) ($M_{anthropomorphic}$ = 0.43, SD = 2.42 vs. $M_{non-nthropomorphic}$ = -0.27, SD = 2.63, F (1, 449) = 4.76, p = .03). Thus, H3 is supported.

Desire for Avoidance. A two-way ANOVA indicated an interaction effect of anthropomorphism of the agent and failure type (F (1,449) = 9.98, p = .002). Further, replicating previous findings with an alternate dependent variable, planned contrasts indicated that for performance failures attributed to anthropomorphic agents (vs. non- anthropomorphic agents), participants exhibited less desire for avoidance, and this effect was marginally significant (H1) ($M_{anthropomorphic}$ = 5.12, SD = 1.41 vs. $M_{non-anthropomorphic}$ = 5.44, SD = 1.26, F(1, 449) = 2.93, p = .09) (see *Figure 10*). Further, in line with previous findings, participants expressed more desire for avoidance for anthropomorphic agents (vs. non- anthropomorphic agents) in the case of for benevolence failures ($M_{anthropomorphic}$ = 5.60, SD = 1.42 vs. $M_{non-anthropomorphic}$ = 5.07, SD = 1.59, F (1, 449) = 7.51, p = .006), supporting H2.

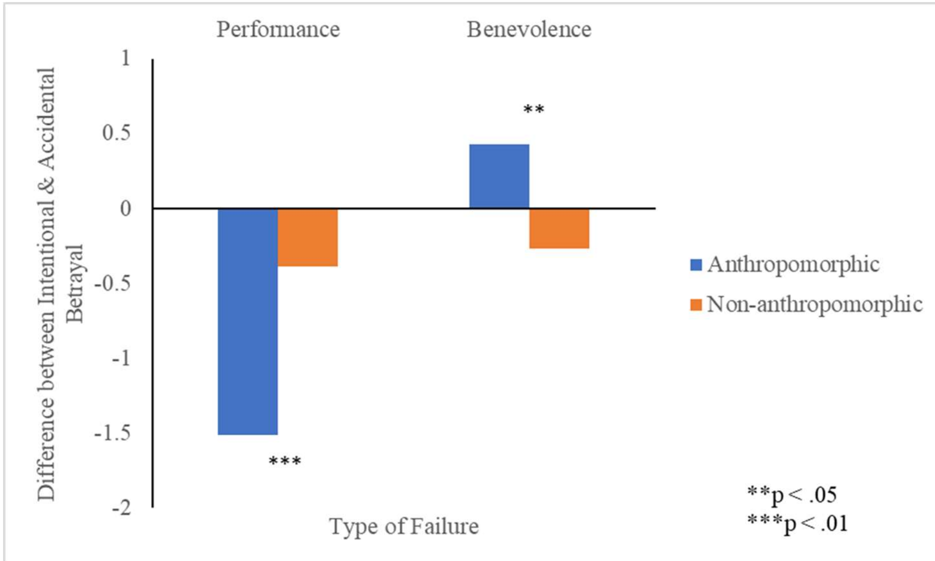


Figure 9. Interaction of Anthropomorphism and Failure Type on Betrayal Perceptions

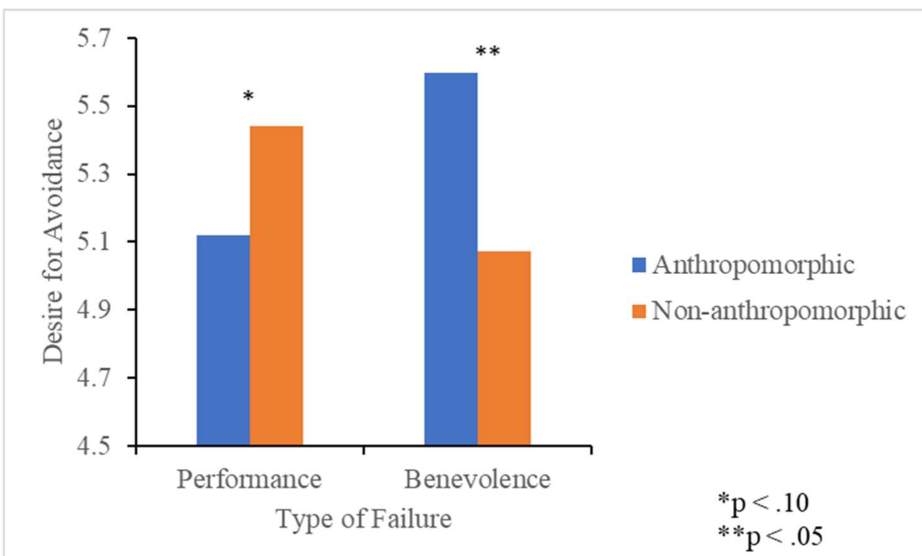
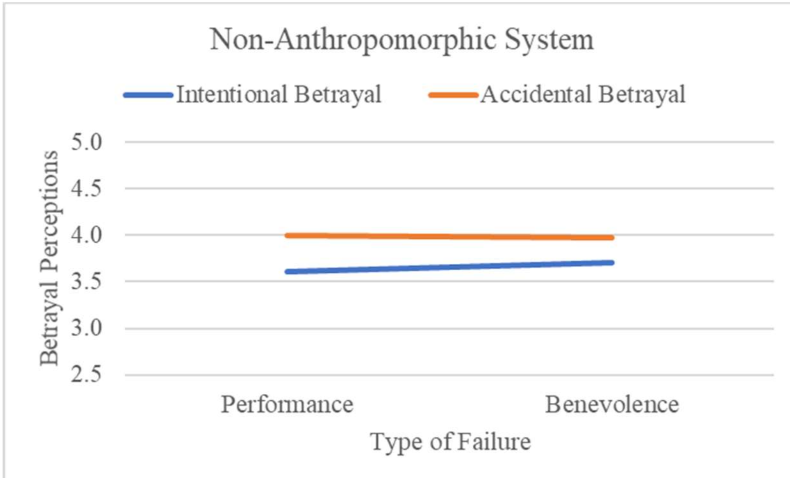


Figure 10. Interaction of Anthropomorphism and Failure Type on Desire for Avoidance

Mediation Analyses. We tested the hypothesized mediating effect of betrayal perceptions on desire for avoidance (H4a and H4b) by using bootstrapping procedures (Preacher & Hayes, 2008). Model 8 was used to test this moderated mediation (Hayes, 2013), such that the independent variable of anthropomorphism of the agent was dummy coded as – 0 = non-anthropomorphic and 1 = anthropomorphic and type of failure (moderator) was dummy coded as – 0 = performance and 1 = benevolence. The differential scores of betrayal perceptions, (as noted earlier) served as the mediator. There was a significant interaction effect of anthropomorphism and failure type on betrayal perceptions (B = 1.82, SE = 0.44; 95% CI = [.94; 2.69]; p = .000). Further, for performance failure, anthropomorphism led to increased perceptions of accidental betrayal (B = −1.12, SE = .31; 95% CI = [−1.73; − .51; p = .000). As hypothesized, this effect reversed for benevolence failures, such that anthropomorphism led to higher perceptions of intentional betrayal (B = .69, SE = .32; 95% CI = [.07; 1.32; p = .03). Further, betrayal perceptions had a significant effect on desire for avoidance (B = .29, SE = .03; 95% CI = [.24; .34; p = .02). The indirect effect of anthropomorphism on desire for avoidance was mediated by betrayal perceptions for both the type of failures, but in opposite direction. Specifically, for performance failures, anthropomorphism led to lower desire for avoidance through betrayal perceptions (B = − .32, 95% CI = [− .51; – .16]. On the other hand, for performance failures, anthropomorphism led to higher desire for avoidance through betrayal perceptions (B = .20, 95% CI = [−.26; –.80]. The direct effect of anthropomorphism on desire for avoidance was not significant for both the failures, indicating a full mediation (for performance failure B = .003, 95% CI = [−.32; .33]; for benevolence failures, B = .33, 95% CI = [−.01; .66].



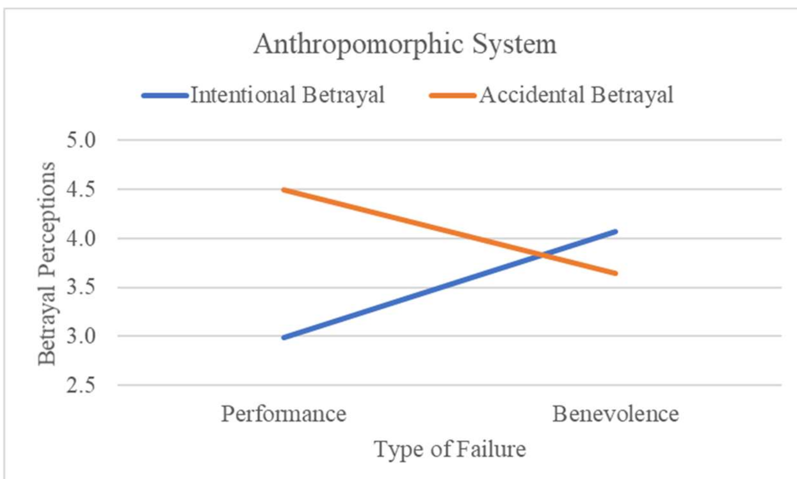


Figure 11. Three-way interaction between Anthropomorphism, Failure Type and Betrayal Type

Additional Analyses. As seen in *Figure 11*, we conducted an alternate form of analyses, a repeated-measures ANOVA with measures for intentional betrayal and accidental betrayal as within-subjects factors and anthropomorphism and failure type as fixed factors was carried out. The ANOVA indicated a significant interaction effect ($F(1, 449) = 16.71$, $p = .000$). Planned contrasts indicated that for performance failures, perceptions of accidental betrayal were higher for anthropomorphic (vs. non-anthropomorphic) agents ($M_{anthropomorphic} = 4.50$, $SD = 1.16$ vs. $M_{non\text{-}anthropomorphic} = 4.00$, $SD = 1.16$, $F(1, 449) = 9.72$, $p = .002$). On the other hand, perceptions of intentional betrayal were lower for anthropomorphic (vs. non-anthropomorphic) in this case ($M_{anthropomorphic} = 2.99$, $SD = 1.49$ vs. $M_{non\text{-}anthropomorphic} = 3.61$, $SD = 1.54$, $F(1, 449) = 9.44$, $p = .002$). Further, in line with previous findings of this study, the results were opposite directionally for benevolence failures. Specifically, for benevolence failures, perceptions of accidental betrayal were lower for anthropomorphic (vs. non-anthropomorphic) agents ($M_{anthropomorphic} = 3.64$, $SD = 1.27$ vs. $M_{non\text{-}anthropomorphic} = 3.97$, $SD = 1.29$, $F(1, 449) = 3.91$, $p = .049$). Lastly, for these failures, perceptions of intentional betrayal were marginally higher for failures attributed to anthropomorphic (vs. non-anthropomorphic) ($M_{anthropomorphic} = 4.07$, $SD = 1.54$ vs. $M_{non\text{-}anthropomorphic} = 3.71$, $SD = 1.62$, $F(1, 449) = 3.13$, $p = .08$).

Discussion

Overall, results for Study 3 indicate that when anthropomorphic agents fail, consumers treat these failures as they would treat failures attributed to humans. Specifically, while performance-oriented failures attributed to such humanlike agents are rated as more of an accidental betrayal (vs. intentional betrayal), benevolence failures committed by such agents are viewed as acts of intentional betrayal. This study further replicated the findings of Study 2, using a more downstream variable: consumers' desire to avoid the system, once it has failed them.

3.6 Study 4

The primary objective of Study 4 was to examine an alternate, and more extreme form of consumers' negative responses toward the failure – consumers' desire for revenge against the services who failed them. Past research on transgressions has noted that willful violations of by services/brands often lead to extremely detrimental consumer behaviors that consumers enact as a reaction to such transgressions (Grégoire, Laufer, and Tripp 2010; Kähr et al. 2016). For instance, Grégoire et al. (2010) show that when consumers perceive firms' transgressions as an intentional act of greed, they engage in revenge behaviors against such firms, in an attempt to avenge these transgressions. Similar findings were noted by Kähr et al. (2016), who note that consumers who believe they have been wronged by the firm exhibit aggressive hostile behaviors such as sabotaging the firm. Thus, given that benevolence-failures attributed to anthropomorphic AI-agents will be viewed as (more) intentional acts of norm violation, we suggest that consumers will exhibit increased desire for revenge against such services. Desire for revenge defined as "consumers' need to punish and cause harm to firms for the damages they have caused" (Grégoire et al. 2009, p. 19), is an emotivational goal which results in several negative consumer behaviors such as spreading negative word of mouth against the firm, vindictively complaining against the firm, and in some cases using physical aggression against the firm (Grégoire, Laufer, and Tripp 2010; Grégoire et al. 2018).

Method

Two hundred and twenty-seven (227) participants were recruited online to participate in a 2 (failure type: performance-based vs. benevolence-based) X 2 (anthropomorphism: yes vs. no) between-subjects design study. Study 4 procedures were similar to study 3, except two important modifications. First, for manipulating anthropomorphism, procedures from Wen Wan et al. (2017) were followed for textual framing. Specifically, in the anthropomorphic condition, the recommendation system was named Skyler and was introduced with human-like descriptions written in first-person language. On the other hand, in the non-anthropomorphic condition, the system was described in machine-like terms using third-person language. Similar to Study 3's procedures, participants were asked to imagine that the recommendation system then asked them their preferences on various characteristics of the laptop, following which, they selected one of the options recommended by the agent. The failure was also manipulated using the same stimuli as study 3. Second, for the current study, participants then indicated their desire for revenge against the firm using a scale adapted from Grégoire et al. (2010) (Cronbach's alpha = 95.4%) using a seven-point Likert scale (see Appendix A for items).

Results

Pretest. The scenario was pretested with one hundred and seventy-four participants recruited online via MTurk. A one-way ANOVA with anthropomorphism as the dependent variable indicated that the participants perceived the anthropomorphic agent as more human ($M_{anthropomorphic}$ = 3.78, vs.

$M_{non\text{-}anthropomorphic}$ = 3.20, F(1,172) = 5.03, p = .026). Further, the pretest indicated that as compared to performance failures, benevolence failures were rated significantly higher on a scale that measured if the system had failed to put their interests above its own. Specifically, ($M_{benevolence}$ = 5.50 vs. $M_{performance}$ = 4.66, F(1,172) = 15.82, p = .000).

Main results. A two-way ANOVA with desire for revenge as the dependent variable indicated a significant interaction effect of anthropomorphism and failure type on desire for revenge (F (1, 223) = 3.94, p = .048). Supporting H2a, planned contrasts indicated that in the case of benevolence-based failures, participants' desire for revenge was significantly higher when such failures were attributed to anthropomorphic agents (vs. non- anthropomorphic agents) ($M_{anthropomorphic}$ = 4.69, SD = 1.54 vs. $M_{non\text{-}anthropomorphic}$ = 3.73, SD = 1.80, F(1, 223) = 8.08, p = .005). On the other hand, no difference in desire for revenge was observed for performance-based failures ($M_{anthropomorphic}$ = 4.17, SD = 1.74 vs. $M_{non\text{-}anthropomorphic}$ = 4.13, SD = 1.81, F(1, 223) = 0.02, p = NS).

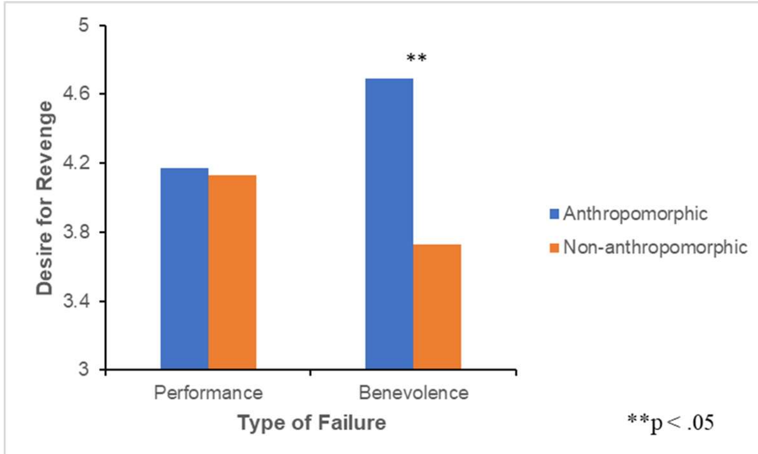


Figure 12. Interaction of Anthropomorphism and Failure Type on Desire for Revenge

Discussion

Taken together, the results of Study 4 highlight the importance of differentiating the two types of failures, given that differences in consumers' desire to avenge such failures persist only in benevolence failures. Given that consumers perceive increased moral violation by anthropomorphic agents (vs. non-anthropomorphic agents) in benevolence-based failures, these results align with recent findings by researchers who show that anthropomorphized technology is perceived to have a higher mind perception of agency (Srinivasan & Abi, 2021). Thus, consumers experience an increased desire for revenge towards such anthropomorphic agents (vs. non- anthropomorphic agents) only when such agents commit benevolence-based failures.

3.7 General Discussion and Conclusion

Across four studies, one including consumers' real-world tweets and another that measured consumers' real-time emotions to real-world failures using an automated facial expression analysis (FaceReader), we show that consumers show attenuated negative responses to performance failures attributed to anthropomorphized (vs. non-anthropomorphized) AI-agents. However, for benevolence failures, consumers exhibit increased negative responses, such as increased desire for revenge, towards anthropomorphic (vs. non-anthropomorphic) agents. A further investigation of the underlying process shows that consumers evaluate benevolence failures committed by anthropomorphic agents as intentional (vs. accidental) betrayal of their trust. On the other hand, performance failures attributed to such anthropomorphic agents are evaluated accidental (vs. intentional) betrayals, indicating that consumers apply the rules of deontic justice for assessing failures committed by anthropomorphic agents.

Theoretical Contributions

The current research makes significant contributions to the understanding of AI-agents. First, we contribute to the scant literature on failures attributed to AI-agents (Dietvorst et al., 2015; Gill, 2020). While a majority of literature in the stream of AI and technology focuses on the adoption of such agents (Longoni and Cian 2020; Gai and Klesse 2019; Leung, Paolacci, and Puntoni 2018), the current understanding about failures due to such agents remains obscure. Given the increasing adoption of AI-agents in every field of life, failures and errors due to such agents are inevitable. Thus, the current research answers the call for research in this area (Khamitov et al., 2020).

Secondly, the current research establishes a typology of failures in the realm of AI. Similar to human errors, task-based AI failures that show a lack of capability on the AI's part, and are classified as performance failures. On the other hand, failures wherein consumers perceive that the system intentionally failed them to put its own interests above theirs is categorized as benevolence-based failures. The current research strongly suggests that consumers react to AI failures varyingly, based on the type of failure they assess it to be.

Further, the current research also contributes to a better understanding about the repercussions of anthropomorphizing AI-agents. Using the context of service failures, our research shows that consumers use the same rules of judgments for anthropomorphic agents, as they would for humans. Specifically, by demonstrating that consumers apply the principles of deontic justice (Folger, Cropanzano, and Goldman 2005; Cropanzano, Goldman, and Folger 2003) for failures attributed to anthropomorphic AI-agents, the current research highlights the circumstances when anthropomorphization of AI-agents attenuates versus exacerbates negative consumers responses in the face of failures. In doing so, we also extend the existing research on anthropomorphization. While literature has extensively investigated the phenomenon of anthropomorphization in the context of branding (Aggarwal & McGill, 2012; Puzakova & Aggarwal, 2018; Puzakova & Kwak, 2017), the current research contributes to a better understanding of this phenomenon in the context of AI and technology.

Managerial Implications

The current research provides significant managerial implications. First, by showing that consumers react varyingly to personified AI-failures based on the type of failure, the current research highlights that anthropomorphizing AI-powered technologies may beneficial for some contexts, but not for others. For instance, in the context of online retail, where consumers often feel that systems such as recommendation engines are rigged in order to benefit the firm, anthropomorphizing the agent may lead to worse consumer outcomes in case of failures. As an example, Hotels.com's FAQ section reveals that one of the factors in the determination of their sort order of recommendations is the compensation paid to this aggregator by the property. Thus, anthropomorphizing systems in contexts where the company has an incentive to deceive the consumer may lead to detrimental outcomes in case of dissatisfactory experiences.

On the other hand, the current research identifies that anthropomorphizing AI-agents works in the favor of firms for a majority of failures, which generally tend to be performance-based failures, such as the agent not being able to interpret the consumers' language, or when such agents misinterpret a consumer. Consumers tend to show decreased negative responses for such failures when they are attributed to anthropomorphic agents, as they would in case of human failures that result because of lack of competence. Thus, the current research clearly identifies conditions for managers, wherein anthropomorphizing such agents would be beneficial versus detrimental in case of service failures.

Lastly, we also provide implications in terms of service recovery. By showing that consumers tend to react more negatively to benevolence failures committed by anthropomorphic agents, the current research suggests that such failures should be prioritized for recovery. Given that firms have limited resources (human, time, etc.), the findings from current research can help managers prioritize their customer service operations more efficiently.

Limitation and Future Research

One of the limitations of the current research is that only two levels of anthropomorphism were used. Given that a technology may be anthropomorphized using varying personification elements, resulting in various levels of anthropomorphization, the current research is limited. For instance, Kim et al. (2019) note that consumers may exhibit negative attitudes towards "too humanlike" robots. Future research could investigate if the results of the current study persist at such high levels of anthropomorphism.

The current research also limits the types of failures to two categories: performance and benevolence. There may be cases, where the failure may lie at the intersection of these two failures, resulting in varying consumer perceptions. The current research also limited itself to failures based on two dimensions of trust: competence and benevolence. There also exists a third dimension of trust in the context of technological agents – integrity (Benbasat & Wang, 2005). Given that integral trust in technological agents builds over time, this element was not considered in the current research. Future research could extend the current research to include a more diverse set of failures.

Lastly, future research could extend the current research by investigating the role of recovery with the interaction of anthropomorphism and failure type. Given that a firm could take different recovery mechanisms such as issuing an apology, compensation or even denial of the failure, this investigation may be fruitful.

References

Accenture. (2019, January 8). *Digital consumer report 2019*. https://www.accenture.com/ca-en/insights/high-tech/reshape-relevance

Aggarwal, P., & McGill, A. L. (2007). Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. *Journal of Consumer Research*, *34*(4), 468–479.

Aggarwal, P., & McGill, A. L. (2012). When brands seem human, do humans act like brands? Automatic behavioral priming effects of brand anthropomorphism. *Journal of Consumer Research*, *39*(2), 307–323.

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 4.

Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing*.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809–825.

Choi, S., Mattila, A. S., & Bolton, L. E. (2020). To Err Is Human (-oid): How Do Consumers React to Robot Service Failure and Recovery? *Journal of Service Research*, 1094670520978798.

Cropanzano, R., Goldman, B., & Folger, R. (2003). Deontic justice: The role of moral principles in workplace fairness. *Journal of Organizational Behavior*, 1019–1024.

Cropanzano, R. S., Massaro, S., & Becker, W. J. (2017). Deontic justice and organizational neuroscience. *Journal of Business Ethics*, *144*(4), 733–754.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

Elangovan, A. R., & Shapiro, D. L. (1998). Betrayal of trust in organizations. *Academy of Management Review*, *23*(3), 547–566.

Fan, A., Wu, L. (Laurie), Miao, L., & Mattila, A. S. (2020). When does technology anthropomorphism help alleviate customer dissatisfaction after a service failure? – The moderating role of consumer technology self-efficacy and interdependent self-construal. *Journal of Hospitality Marketing & Management*, *29*(3), 269–290. http://10.0.4.56/19368623.2019.1639095

Fan, A., Wu, L. L., & Mattila, A. S. (2016). Does anthropomorphism influence customers' switching intentions in the self-service technology failure context? *Journal of Services Marketing*.

Folger, R., Cropanzano, R., & Goldman, B. (2005). What is the relationship between justice and morality. *Handbook of Organizational Justice*, *215*, 215.

Folkes, V. S. (1988). Recent attribution research in consumer behavior: A review and new directions. *Journal of Consumer Research*, *14*(4), 548–565.

Folkman, S., & Lazarus, R. S. (1984). *Stress, appraisal, and coping*. New York: Springer Publishing Company.

Gai, P. J., & Klesse, A.-K. (2019). Making recommendations more effective through framings: impacts of user-versus item-based framings on recommendation click-throughs. *Journal of Marketing*, *83*(6), 61–75.

Gill, T. (2020). Blame it on the self-driving car: how autonomous vehicles can alter consumer morality. *Journal of Consumer Research*, *47*(2), 272–291.

Grégoire, Y., Laufer, D., & Tripp, T. M. (2010). A comprehensive model of customer direct and indirect revenge: Understanding the effects of perceived greed and customer power. *Journal of the Academy of Marketing Science*, *38*(6), 738–758.

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, *60*(11), 2169–2188.

Kähr, A., Nyffenegger, B., Krohmer, H., & Hoyer, W. D. (2016). When hostile consumers wreak havoc on your brand: The phenomenon of consumer brand sabotage. *Journal of Marketing*, *80*(3), 25–41.

Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, *62*(1), 15–25.

Khamitov, M., Grégoire, Y., & Suri, A. (2020). A systematic review of brand transgression, service failure recovery and product-harm crisis: integration and guiding insights. *Journal of the Academy of Marketing Science*, 1–24.

Kim, S. Y., Schmitt, B. H., & Thalmann, N. M. (2019). Eliza in the uncanny valley: anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing Letters*, *30*(1), 1–12.

Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS Quarterly*, 941–960.

Kramer, T., Spolter-Weisfeld, S., & Thakkar, M. (2007). The effect of cultural orientation on consumer responses to personalization. *Marketing Science*, *26*(2), 246–258.

Leung, E., Paolacci, G., & Puntoni, S. (2018). Man versus machine: Resisting automation in identity-based consumer behavior. *Journal of Marketing Research*, *55*(6), 818–831.

Li, D., Rau, P. L. P., & Li, Y. (2010). A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics*, *2*(2), 175–186.

Lin, Y.-T., Doong, H.-S., & Eisingerich, A. B. (2020). Avatar Design of Virtual Salespeople: Mitigation of Recommendation Conflicts. *Journal of Service Research*, 1094670520964872.

Longoni, C., & Cian, L. (2020). Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The “Word-of-Machine” Effect. *Journal of Marketing*.

Lopatovska, I., & Williams, H. (2018). Personification of the Amazon Alexa: BFF or a mindless companion. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, 265–268.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535–556.

Mourey, J. A., Olson, J. G., & Yoon, C. (2017). Products as pals: Engaging with anthropomorphic products mitigates the effects of social exclusion. *Journal of Consumer Research*, *44*(2), 414–431.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). " Alexa is my new BFF" Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2853–2859.

Puzakova, M., & Aggarwal, P. (2018). Brands as rivals: Consumer pursuit of distinctiveness and the role of brand anthropomorphism. *Journal of Consumer Research*, *45*(4), 869–888.

Puzakova, M., & Kwak, H. (2017). Should anthropomorphized brands engage customers? The impact of social crowding on brand preferences. *Journal of Marketing*, *81*(6), 99–115.

Srinivasan, R., & Abi, G. S. (2021). When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors. *Journal of Marketing*.

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, *20*(1), 43–58.

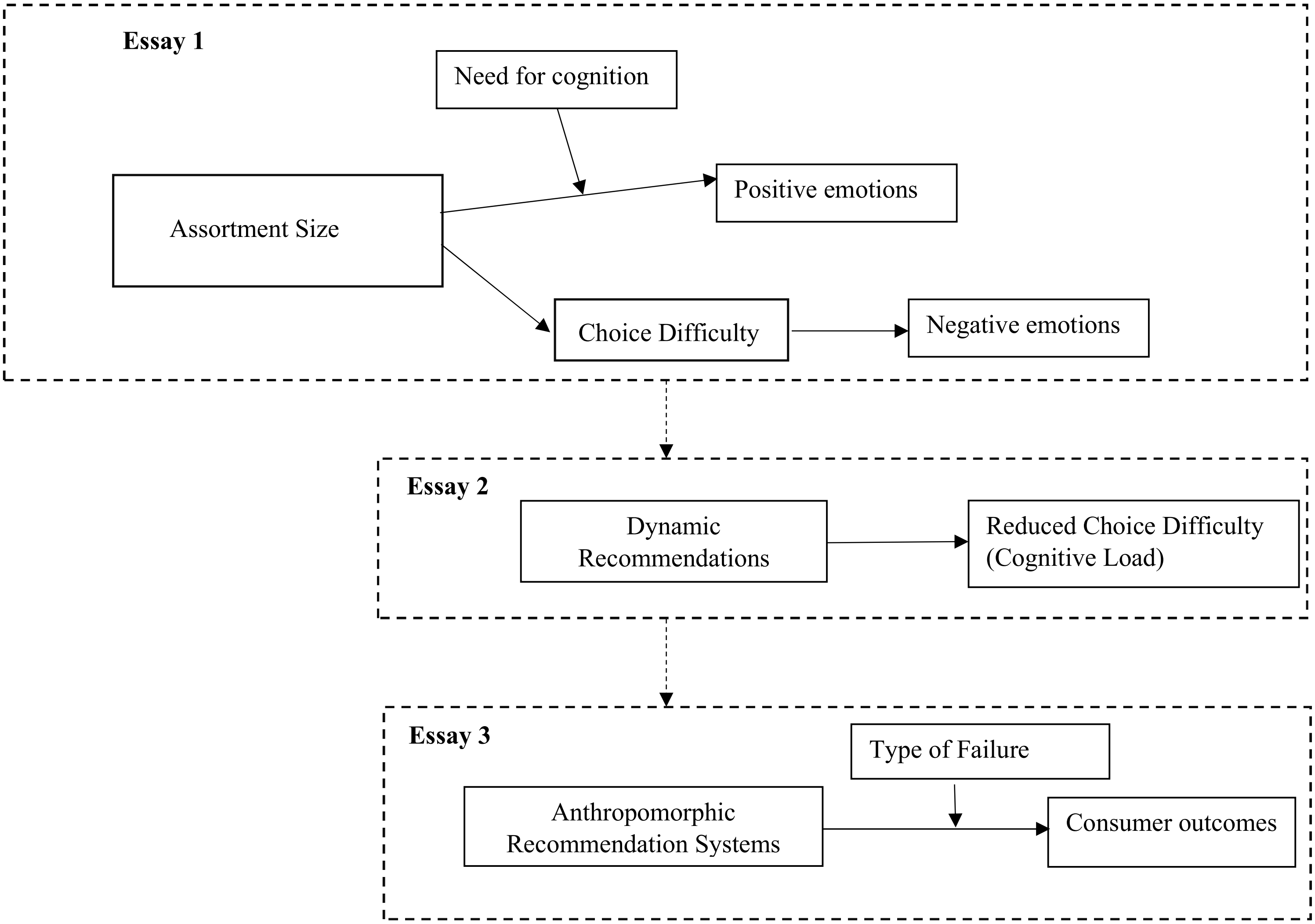
Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117.

Wen Wan, E., Peng Chen, R., & Jin, L. (2017). Judging a book by its cover? The effect of anthropomorphism on product attribute processing and consumer preference. *Journal of Consumer Research*, *43*(6), 1008–1030.

Conclusion



Overall Conceptual Framework of the Thesis

Taken together, my theses accomplish three objectives. The first goal of the current research was to investigate and contribute to the debate on choice overload., by employing psychophysiological measures. By doing so, we attempt to reconcile the previous findings in this context and to lead to a better understanding of the underlying process. Overall, our findings indicate the existence of increased mixed emotions, that is, a simultaneous increase in positive and negative emotions as a function of assortment size. These findings help understand the findings of a previous meta-analyses which found a null main effect of choice, since an increase in choice essentially leads to both – positive and negative outcomes among consumers. Further, the results show that while choice difficulty mediates the effect of assortment on negative emotions, consumers' need for cognition moderates the effect on positive emotions.

Overall, the results of Essay 1 not only contribute theoretically to the existing literature on choice overload, but also lead to important managerial implications. For instance, our findings of mixed emotions as a function of assortment size imply that reducing options as a strategy for reducing choice overload – a mechanism that leading businesses such as Tesco are adopting to simplify shopping (Wood & Butler, 2015) – may backfire. Although reducing assortment would lead to decreased choice difficulty (and hence to a resulting decrease in negative emotions associated with increased choice), it would also devoid consumers of experiencing increased positive emotions. Consumers, especially ones with higher need for cognition, seek variety while making purchase decisions. A larger assortment makes purchase decisions enjoyable for such consumers, as these individuals are intrinsically motivated to derive positive outcomes – the probability of which is greater with a larger assortment size. Thus, these findings indicate that businesses need to adopt strategies such would reduce choice difficulty, without essentially reducing the assortment size – a solution that could be accomplished by using strategies such as deployment of dynamic, context-aware recommendation systems that could help consumers ease the decision-making process, without reducing the size of assortment per se.

Building on the findings of Essay 1, Essay 2 implements and tests a novel, adaptive personalized recommender system that aims to ease consumer decision-making process. By implementing and testing this novel recommender system that is able to assess consumers' cognitive load in real-time through neurophysiological tools, this essay contributes to the literature of adaptive personalized systems (Chung et al., 2009, 2016; Hauser et al., 2009; Urban et al., 2014) and paves way for futuristic systems that will be able to adapt themselves based on consumers' implicit, real-time neurophysiological measures. This dynamic system, which is able to present recommendations at an appropriate time when the consumers' need its help, also helps shed light on a theoretical debate that such dynamic systems could lead to: whether such dynamic systems, that present recommendations to consumers after the consumer has already begun her/his decision-making process help, or instead hinder this decision-making process. Specifically, this system tests two competing hypotheses that such dynamic recommendations could lead to: alleviation of choice difficulty due to a reduction of consideration set or the increment of choice difficulty due to enlargement of the consideration set. Findings from this essay, which suggest a reduction in cognitive load (measured using neurophysiological tool) due to dynamic recommendations indicate that such systems could help ease consumer decision-making process by easing the choice difficulty. Given that Essay 1 shows that consumers do experience increased choice difficulty with an increase in assortment size, the findings of Essay 2 provide a possible solution for reducing this difficulty, without reducing the assortment size itself.

Essay 2 further tests two alternative framings that such dynamic recommender systems could use, and the findings suggest that recommendations that are framed in terms of "other similar consumers" lead to improved consumer outcomes, as compared to recommendations that are framed as "personalized recommendations from the system". Taken together, Essay 2 helps guide practice on the deployment of such futuristic recommender systems, which could be personalized for each consumer, based on their individualistic cognitive needs. Given the benefits of such dynamic recommender systems, managers could adapt the recommender system used in this essay to employ other implicit consumer inputs, such as consumers' real-time clickstream data, and present dynamic recommendations.

Finally, given the findings of Essay 2, which show that recommender systems offer consumer benefits, Essay 3 explores how would consumers react when they seek the help of a recommendation system, and it ends up failing them. Given that many real-world virtual agents such as recommender systems, chatbots, etc. are increasingly being anthropomorphized using personification elements such as name, gender, appearance, etc., this essay investigates the role of such anthropomorphized recommendation agents in the context of failures. Overall, the findings of this essay show that consumers experience less negative reactions when an anthropomorphized (vs. non- anthropomorphized) system commits a performance-based failure, which are more common in nature. Further, this essay also identifies the conditions where such anthropomorphic systems fare worse – that is benevolence failure, that is, failures wherein the technological agent fails to care about the consumers' interests. Overall, by distinguishing between the types of failures, this essay highlights conditions and contexts wherein anthropomorphized systems fare better (vs. worse) than their non-anthropomorphized versions.

Our findings indicate that in line with previous research that has shown benefits such as increased adoption and engagement due to anthropomorphization of technology, such anthropomorphized agents also fare better in a majority of real-world failures, which happen to be performance-oriented in nature. These findings help mangers identify contexts where anthropomorphization of the agent would help versus hurt them in case of failures. For instance, not long ago, Microsoft found itself at the heart of a major crises due the failure of its AI chatbot Tay (Vincent, 2016). Tay, a technological agent which was anthropomorphized using a name, gender and humanlike physical attributes, generated a lot of negative publicity when it gave the consumers the impression that it did not care about their interests – a failure which would be categorized as a benevolence-based failure as per our theorizing. Had the same failure occurred with a non- anthropomorphized AI chatbot, our research suggests that the negative repercussions would have been milder, highlighting the importance of the current research.

Appendix A

Measure	Used in….
Anthropomorphism It seems almost as if the recommendation system is like a person	Pretests for Studies 1, 2, 3 and 4
Benevolence Please rate how much do you agree with the following statements 1. The recommendation system did not put my interests first 2. The recommendation system did not keep my interests in mind 3. The recommendation system placed its self-interests above mine	Pretests for Study 2 and Study 4
Negative Affect Given your experience, please describe how do you feel about your experience I feel…. 1. Negative (1= Not at all, 2= A little, 3 = Moderately, 4 = Quite a bit, 5 = Extremely) 2. Hostile (1= Not at all, 2= A little, 3 = Moderately, 4 = Quite a bit, 5 = Extremely)	Study 2

Desire for Avoidance Please rate how much do you agree with the following statements. I want to... 1. Keep as much distance as possible between the recommendation system and me. 2. Avoid using the recommendation system 3. Cut off the relationship with the recommendation system	Study 3
Intentional Betrayal Please rate how much do you agree with the following statements. It believe that... 1. The recommendation system intended to betray me 2. The recommendation system willingly cheated me 3. The recommendation system took advantage of me	Study 3
Accidental Betrayal Please rate how much do you agree with the following statements.	Study 3

It believe that... 1. The recommendation system accidentally erred 2. The recommendation system had good intentions 3. The recommendation system did not intend to let me down	
Desire for Revenge Please rate how much do you agree with the following statements. 1. I want to take actions to get the online retail company in trouble. 2. I want to cause inconvenience to the online retail company. 3. I want to punish the online retailer in some way. 4. I want to make the online retailer get what they deserve. 5. I want to get even with the online retailer company.	Study 4

Appendix B

Additional Study 1B: Alternative operationalization Using the context of recommendation systems

For an alternative operationalization of Study 1, an additional study was conducted using the context of recommendation systems. Given that studies 2 and 3 were conducted using the context of AI-based recommendation agents, this context was selected for Study 1B as well.

Method

Tweets regarding failures of recommendation agents were scrapped using procedures similar to Study 1. The scraping of Twitter resulted in 227 tweets, which were scrutinized by a research assistant to identify only those tweets that indicated a failure experienced by consumers due to a technological recommendation system (example of excluded cases were tweets that were recommendations by a person as opposed to a recommendation agent). This data cleaning resulted in data set of 131 tweets (data is available upon request). The following review is one of the examples included in the final dataset : "Booked a hotel through @hotwire to spend New Years in Madrid , I continue to get emails for offers in Madrid #recommendationfail #annoying". Recent literature has shown that anthropomorphism is induced by personification, such that consumers often attribute personified pronouns as well as humanlike qualities to anthropomorphic agents (Purington et al. 2017; Lopatovska and Williams 2018). Using personification by consumers as a proxy for anthropomorphism of agents, the research assistant was asked to code tweets wherein the system was personified as "anthropomorphic" (vs. non-anthropomorphic). For instance, the tweet "Seriously, Twitter... Even after all of my #Mets tweets, you suggest I follow the #Marlins!? #RecommendationFail" was coded as anthropomorphic. The research assistant was also asked to classify these tweets into performance-based failures and benevolence-based failures. However, all the retrieved tweets were identified as performance failures, as opposed to benevolence failures, indicating that performance oriented failures are more abundant in real life, as seen in Study 1 as well.

Similar to study 1, these tweets were analyzed using LIWC, and measures for tone of the tweet were collected.

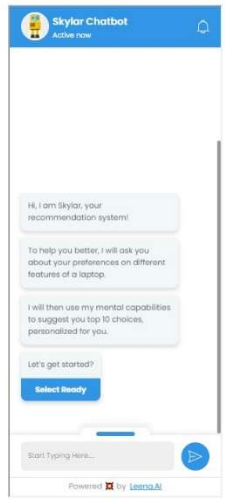
Results and discussion

Similar to Study 1, an ANOVA with emotional tone as the dependent variable and anthropomorphism of the agent as the independent variable indicated a significant effect ($F(1, 129) = 7.10$, $p = .009$). Further, the tweets coded as those attributed to anthropomorphic (vs. non- anthropomorphic) systems were significantly more positive in tone ($M_{anthropomorphic} = 55.38$, $SD = 41.29$ vs. $M_{non\text{-}anthropomorphic} = 39.32$, $SD = 33.74$).

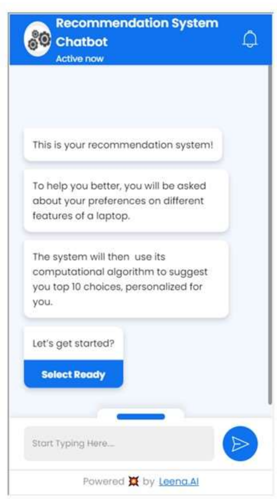
Similar to Study 1, this alternative study indicated that consumers exhibit decreased negative responses when they attribute performance failures to anthropomorphic (vs. non-anthropomorphic) agents. Using an alternative context and mechanism for coding anthropomorphism of AI-agents, this study replicates the findings of study 1, further validating the robustness of previous findings.

Stimuli for Study 2

Anthropomorphic Version of Recommendation Chatbot



Non-anthropomorphic Version of Recommendation Chatbot



Set of laptops suggested by the Recommendation Agent

HP - ENVY 17.3" Touch-Screen Laptop - $1,249.99
Intel Core i7 - 12GB RAM - 6 pounds

HP ENVY - 13.3" Touch-Screen Laptop - $834.99
AMD Ryzen 7 4700U - 8GB RAM - 3.1 pounds

ASUS - ROG Zephyrus S15 15.6" - $2,999.99
Intel Core i7 - 32GB RAM - 4.19 pounds

Dell - XPS 2-in-1 13.4" - $1,599.99
Intel Core i7 - 16GB RAM - 2.89 pounds

Microsoft - Surface Laptop 3 - 15" Touch-Screen - $2,799.00
AMD Ryzen 7 - 32GB RAM - 3.4 pounds

HP ENVY 15.6" Touch Laptop - $1,349.99
Intel Core i7-10750H - 16GB RAM - 3.2 pounds

MSI - Creator 15 - 15.6" 4K UHD - $2,599.99
i7-10875H - 32GB RAM - 5 pounds

Lenovo - 14" ThinkPad - $1,199.00
AMD Ryzen 5 PRO - 8 GB RAM - 3 pounds

Samsung - Galaxy Book 15.6" Touch-Screen Laptop - $1199.99
Intel Core i7 - 12GB RAM - 3 pounds

Lenovo - 13" ThinkBook - $864.99
Intel Core i5 - 8GB RAM - 3 pounds

List of Top "5" laptops

Dell XPS 13 – 13.4" - $ 999.99
Intel Core i3-1005G1 – 4 GB RAM – 2.9 pounds

Acer Chromebook Spin – 13.5" - $629
Intel Core i5 – 8 GB RAM – 3.20 pounds

Asus ZenBook 13 UX333FA - 13.3" – $999
Intel Core i5 – 8 GB RAM – 2.7 pounds

Dell Inspiron - 13" – $ 899.99
Intel Core i7 – 16 GB RAM – 3.2 pounds

Microsoft Surface Pro- 12.3" – $ 959 Touch-Screen
Intel Core i3 – 4 GB RAM – 2.8 pounds

Stimuli for Study 3

Anthropomorphic Version

Non-anthropomorphic version