ADDRESSING THE RECOMMENDER SYSTEM DATA SOLICITATION PROBLEM WITH ENGAGING USER INTERFACES

by

Quang V. Dao

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

December 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Steven J. Landry, Co-Chair
Industrial Engineering

Dr. Denny Yu, Co-Chair
Industrial Engineering

Dr. Brandon Pitts
Industrial Engineering

Dr. Walter W. Johnson
NASA, Retired

Approved by:
Dr. Abhijit Deshmukh

I dedicate this work to my dearest wife Astrid Dao, for her unconditioned support throughout this endeavor, and who carried the full responsibility of caring for our son Hugo to allow me the capacity to see to the completion of this document. I also owe an unrepayable debt to my mentor, Dr. Walter W. Johnson, whose insights inspired the development of the ideas herein, and encouraged me to press forward on these ideas, even when I did not yet see the value in them myself. If our weekly meetings do not continue after I've deposited this document, I will miss our lively and insightful discussions. I am incredibly thankful to Dr. Steven J. Landry for giving this mediocre scholar the opportunity to pursue this degree, and for teaching me the value of an unwavering pursuit of the truth. Also, by way of this Ph.D. program, paved the way for my current and future occupational success. Finally, I thank my parents, Dung Ly and Clifton Dao who supported me through all my years of schooling, and my brother Minh Dao, who picked up my responsibilities at home to also afford me the spare capacity to focus on this effort.

ACKNOWLEDGMENTS

I would like to thank Dr. Joel Lachter, Dr. Jessica Nowinski, and Dr. Alonso Vera for continued support at the NASA Human System Integration Branch. Also, I extend my gratitude to Dr. Sean Laraway for his support with graduate students and participants for conducting my studies.

TABLE OF CONTENTS

LIST OF TABLES 7
LIST OF FIGURES 8
ABSTRACT 9
1. INTRODUCTION 10
1.1 Problem Statement: The Data Solicitation Problem 10
1.2 Contribution 12
1.3 Organization 14
2. Recommender Systems and the Role of Data 16
3. The Context Aware Recommender System (CARS) 19
4. INPUT DATA ACQUISITION FOR RECOMMENDER SYSTEMS 22
4.1 Ratings Data and the Data Solicitation Problem 22
4.2 Beyond Explicit Ratings: Contextual Information 23
4.3 Obtaining Contextual Information 25
5. HUMAN COMPUTER INTERACTION APPROACHES TO ENGAGEMENT 27
5.1 Joint Activity 27
5.2 What is Engagement? 29
5.3 Creating Narratives 43
5.4 Mimicking Human Interaction for Engagement 46
5.4.1 Anthropomorphism 47
5.4.2 Verbal Presentation: Mimicking Human Conversation with Social Discourse 50
5.4.3 Cognitive Presentation: The Illusion of Human Intelligence 54
5.4.4 Visual Presentation: Behavioral and Form Realism 56
5.5 Trust 60
6. THE EMERGENCE OF HUMAN MACHINE TEAMS 66
6.1 Levels of Automation (LOA) 66
6.2 Autonomy 69
6.3 Human Machine Teaming Frameworks: A Few Considerations 70
6.3.1 Theoretical Frameworks 71

6.3.2 Practical Frameworks 74
7. FUTURE DIRECTIONS 82
APPENDIX A. EXPLORATORY STUDY 100
APPENDIX B. PRE-SESSION QUESTIONNAIRE 108
APPENDIX C. ANTHROPOMORPHISM INDEX 114
APPENDIX E. NUMBER OF RATINGS OBTAINED FROM PARTICIPANTS ACROSS CONDITIONS 120
APPENDIX F. PRE-SESSION QUESTIONNAIRE OUTPUT 123
APPENDIX G. ANTHROPOMORPHISM INDEX OUTPUT 125
APPENDIX H. POST SESSION QUESTIONNAIRE OUTPUT 127
APPENDIX I. AGENT BACKSTORIES 128
APPENDIX J. AGENT REQUEST SCRIPT 130
REFERENCES 131

LIST OF TABLES

Table 1. Engagement emerges from an appropriate alignment of activity to motivations. 42

LIST OF FIGURES

Figure 1. The collaborative filtering process. 16

Figure 2. The contextual filtering process. 19

Figure 3. The O'Brien and Tom [14] process model for user engagement. 33

Figure 4. Determinants of engagement: a simplified version of the Perski et al. [13] framework. 39

Figure 5. Basic conversational structure. 51

Figure 6. A framework for trust, joint activity and engagement. 61

Figure 7. Collaborative control model. Adapted from Johnson et al. [44]. 73

Figure 8. Function hierarchy. Adapted from IJtsma et al. [28]. 76

Figure 9. The goal space for a two-agent conversation. Adapted from Wallis [147]. 87

Figure 10. No agent static object. 92

Figure 11. Agent. 92

Figure 12. A conceptual framework on the direct and indirect influences of interface implementation concepts on engagement and user data output. All arches represent existing relationships identified in the empirical work. Dotted arches represent relationships not well supported by data and are prime research areas. 96

ABSTRACT

The motivation for this work comes from the need to obtain data for autonomous systems that rely heavily on recommender systems for human interaction. Recommender systems are data-driven technologies and depend heavily on large quantities of user data to function effectively, e.g. [1], [2]. However, acquiring that data has proven difficult [3]–[7] because users typically do not want to lend the effort to furnish the data. Part of the reason for this problem stems from the reluctance of users to provide data because, as reported by users, it is cognitively taxing and/or too tedious to do so [8], [9]. With autonomous systems bringing greater demand for user data, in some applications, this also brings an opportunity to solicit data from users. The American driver spends about an hour each day on the road; with self-driving cars, this means there will be a captive audience in the vehicle for at least the duration of the trip [10]. To exploit this, a user interface will need to be designed to coax the user into achieving system goals, like data solicitation. One approach is to design a system to leverage an already present tendency for people to socially interact with technology [11]. **In this thesis, I argue that such an approach would involve incorporating interaction concepts that facilitate engagement into the design of recommender system interfaces that will improve the likelihood of obtaining data from users. To support this claim, I synthesize past work on human-computer interaction and recommender systems to derive a framework to guide scientific investigations into interface design concepts that will address the data solicitation problem.** In addition**,** I present the results of a study of how anthropomorphism, as an indicator of engagement [12], may affect the amount of data provided by a user. I begin with a discussion of the problem and then provide a description of the recommender system filtering process to illustrate why user data is important. Then I describe the types of data that will be relevant to recommender system functioning, as the type of data determines how it is obtained. Subsequently, I introduce the construct of engagement and discuss design concepts for interactions that can potentially support it. I conclude with a discussion of future empirical work aimed at testing elements of the human-computer interaction approaches presented herein. The contribution will be the framework stated above and a research agenda stemming from that framework.

1. INTRODUCTION

1.1 Problem Statement: The Data Solicitation Problem

Recommender systems (RS) are becoming more common with applications that range from commercial product selection to route planning for transportation systems; they take a very large set of options and reduce it to a more manageable subset. Preference driven RSs require very large data sets. However, a common issue with state-of-the-art RS interaction techniques is that they do not solicit adequate input from users to generate options acceptable to them; this is the data solicitation problem. **The thesis herein is that incorporating interaction concepts that facilitate engagement into the design of recommender system interfaces will improve the likelihood of obtaining data from users.** Engagement can be defined as a positive experience that users have with other people and technology [13], [14]; this is a simplified definition and I will return to engagement in Chapter 5. For technology, an indication of engagement is repeated use [13], [15]. The presumption that is being made in this work is that enticing users into continued interaction with recommender systems will create opportunities to solicit data from users. However, soliciting engagement alone does not guarantee more user data. Recommenders system interfaces will need to be designed to take advantage of the opportunities provided through engagement to obtain user data. When a system is designed to encourage coordinated action between the user and machine, a concept that will be explored in this work, user data can be obtained as part of the information exchanges that occur. In recommender systems that allow users to control what filters are applied to the options being generated, users implicitly reveal their interests based on what constraints they apply to the solution set. For example, future self-driving car concepts promise to relieve the human from manual operation of the vehicle. These cognitive resources can then be reallocated to responding to rating prompts from the system (e.g., from scale of 1 to 5, rate the acceptability of a route), input contextual constraints, such as limiting routes to those with the fewest pedestrians, or lift constraints on the RS to reveal new options. Those responses can then be fed to the recommender system for generating route options that are sensitive to high-level goals and objectives outside of minimizing time and distance, such as path options that cross the most points of interest, or rerouting to a gym to pass the time that would otherwise be consumed in idle traffic. By negotiating route options through information exchanges with the vehicle RS in the above process, the human would effectively be engaging the system in data solicitation without it being a separate effort. In vehicle operations such as self-driving cars, the recommender system has a captive audience for at least the duration of the trip. If self-driving vehicles are realized, this will be an untapped opportunity to obtain more data from users. However, proposed interface design concepts need to be effective at engaging users to leverage that opportunity. Since information exchange depends so much on communication with the human, designers will need to bring to bear the latest advancements in technology to make communicating and interacting with the machines engaging. The challenge will be in implementation and determining what design concepts will be effective. This work is concerned with identifying what interface concepts show the most promise for positively influencing engagement and coordinated action with machines, and whether these interactions will result in more user data.

The data solicitation problem stems, at least in part, from the reluctance of users to provide data because it is cognitively taxing and/or too tedious to do so [8], [16] and the task would carry next to no priority unless it demonstrates some practical or recreational value. Some systems have been designed which attempt to work around the data solicitation problem, but at the risk of compromising trust and acceptance of the system recommendations [17], [18]. Those systems accept that the initial set of recommendations will probably not be supported by data and are likely to be rejected. They rely on the chance that the user will provide feedback on the initial set and then continue to interact with the system to provide more data. However, this continued interaction is not guaranteed, because the presentation of poor recommendations increases the risk that the system is abandoned before it acquires the data it needs to refine those recommendations.

The data solicitation problem is important because when recommender systems fail to produce acceptable results due to the scarcity of input from users, the overall utility of the product in which they are embedded diminishes. Conversely, when more data is available, recommender systems become increasingly accurate. Recommendation accuracy is the most common measure of the quality of the results generated by recommender systems by developers. Greater ***recommendation accuracy*** is achieved when the difference between the recommender system's prediction of a user's rating for an option and the actual rating provided by the user is small.

Addressing the data solicitation problem amounts to keeping the user engaged so that they do not abandon the system before it has collected enough input from users to serve useful recommendations. The current work aims to examine interaction approaches to this quandary. In particular, the work will introduce a reason, apart from receiving recommendations, for why a user would continue to provide inputs to the system. This reason lies with experiencing engagement with a system. Designing a machine interface for engagement will need to leverage models of human-to-human interaction; what implementation techniques will be most effective may be a function of the application area. In high workload environments, the user will not have the spare capacity to fully engage an RS but natural language interaction might address such issues. When self-driving vehicle technology reaches maturity and advances in artificial intelligence afford automated systems the ability to share cognitive functions with the human, engaging interfaces can leverage aspects of human-human interaction to facilitate cooperation between humans and machines to improve RS output, as well as keep humans operationally involved so that functions can be effectively recovered in the event of automation failure. This way, the human continues to play a meaningful role in the operation of the vehicle as advances in self-driving cars and other automated systems, such unmanned aircraft systems, continue to separate the human from manual operations.

1.2 Contribution

The contribution of this thesis is to provide a framework to guide scientific investigations into interface design concepts that will address the data solicitation problem. Based on this framework, I will suggest an agenda for future empirical work to uncover relationships - if they exist - between human computer interaction techniques and the likelihood that users will volunteer data to the system. In that agenda, two principle assumptions will need to be tested. The first is that more user data leads to more relevant RS options. Although developers have measures to conveniently quantify how closely RS output represents user preference, there is little evidence that the output will be appropriate to the operation, e.g., route options for vehicle operations, or acceptable to intended users. The second assumption is that more joint activity will lead to the quantity and type of information that will be ingestible to the RS. Joint activity is activity that is done in coordination between multiple actors. Due to the coordination aspect of joint activity, communication must occur. The presumption is that if this joint activity is conducted between a machine and a human, the machine would be able to attain user data during the ongoing information exchange required of the joint activity. If these assumptions are validated, continued work can then focus on how to best implement social discourse with machines to support the exchange of information. One important issue related to implementation is determining what rules and common human practices can be quantified and applied to designing social interactions with an RS, or machines in general. These rules, among other goals of communication, help to facilitate managing speech to encourage another actor to divulge more information, like identifying common interests and then asking more follow-on questions about those interests. Another necessary issue to be addressed is how to model the human in a discourse with a machine, that is, how will the human communicate with a machine given their interests, dislikes, and values. Also, what social or cognitive factors must be modeled? What are the mechanisms involved? How the user may respond to the machine that is exercising greater autonomy by proactively engaging the user in social interaction will also need to be studied. Even if a machine can perfectly mimic a human conversationalist, there is no guarantee that a user will reciprocate the attempt by the machine to socially interact. Finally, the best application areas for future work must be determined, along with disadvantages of each based on which will afford the best opportunity for social interactions, e.g., personal assistant applications may be more amenable to studying social interactions than a word processing application. The rest of this document provides theoretical and empirical justification for pursuing this research agenda – summarized below.

Two principle assumptions must be tested:

- Obtaining contextual information leads to more relevant RS recommendations
- More social discourse will lead to more data from users

Future work should address the following for developing engaging recommender system interfaces:

- <u>Model human social interaction</u>: social constructs that apply to development of conversation planning must be determined
- <u>Model the human interlocutor</u>: factors and mechanisms for modeling the users in conversation must be determined
- <u>Evaluate user reaction to novel machine roles</u>: users' response to machines exercising greater autonomy in initiating interactions must be determined

- Identify appropriate context for investigation: the best applications areas for future work must be determined, along with advantages and disadvantages of each based on
    - What applications afford the best opportunity for social interactions?
    - What applications will require recommender systems to be effective?
    - What applications will require information from the human in order to be effective?

1.3 Organization

In Chapter 2 I begin with a description of recommender systems. I describe the underlying mechanisms and techniques that make the RS work. The main purpose of the chapter is to point to where input from the human is important and why. I identify what goals need to be achieved by the recommender system, such as diversity in the options it presents, and how that is driven by data from the human. I will cover two fundamental filtering techniques, content-based filtering and collaborative filtering, because they are often used together; the former to jumpstart the RS process, and the latter being an iterative process to improve recommender accuracy. In Chapter 3, I expand my discussion of recommender systems to include context aware recommender systems (CARS). This type of recommender system is built on top of the filtering techniques from Chapter 2 and incorporates context data from users. Context data serves to provide more relevance to the options generated by the RS and affords the system the ability to respond appropriately to changing operational situations; it highlights the human's crucial role as a provider of context to an operation. Chapter 4 describes the common types of data that drive recommender systems and how they are obtained. I place special emphasis on defining context data. As humans play an important role in provided contextual data, user interfaces will need to entice interaction with the system through engaging experiences to obtain that data. This motivates the engagement approaches in Chapter 5 and 6. I introduce joint activity in Chapter 5 to identify the different types of activities that can occur between actors; this is important because an interface concept can fail to deliver an engaging experience based on the type of activity it is meant to support. In the same Chapter, I define engagement and then I proceed to describe various interaction concepts that can potentially have a positive impact on engagement. Chapter 6 pertains to the topic of human machine teams, which is an extension of joint activity, but specific to how humans will interact with machines that are designed to exercise more proactive behavior. The presumption I am making here is that this proactive behavior will encourage more joint activity and subsequently exchanges of information where more user data can be obtained. In Chapter 7, I close with describing the promised framework and propose a research agenda to address the topic of this thesis; addressing the recommender system data solicitation problem with engaging user interfaces.

2. RECOMMENDER SYSTEMS AND THE ROLE OF DATA

Recommender systems are automated systems that reduce a very large set of options into a smaller more manageable set of recommendations by employing complex algorithms, rules, and heuristics [19]. These recommendations can be personalized or non-personalized [20]. Recommender systems provide personalized recommendations when they incorporate the needs and preferences of the user [21]. When personalization is desired, collaborative filtering is commonly employed. Collaborative filtering is a technique that predicts a user's preference for items based on how similar they are to other users in the community [2], [22]. Popular Internet vendors, such as Amazon.com, who employ collaborative filtering, typically explain their recommendations by stating that "other people who bought item x also bought item y."

The recommendations are poor when recommender systems do not have access to large volumes of data. This is particularly true with systems designed to personalized recommendations. One technique for achieving personalized recommendations is with **collaborative filtering** [2], [23]. In Figure 1 I provide an illustration of the collaborative filtering process. I separate the process into two phases - Phase 1 and Phase 2.

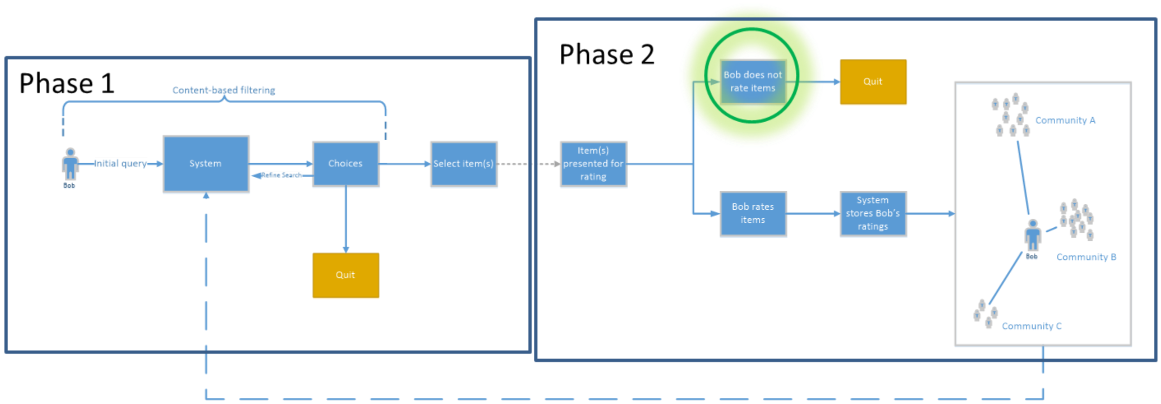


Figure 1. The collaborative filtering process.

If, a user Bob, had visited this system before, the initial query will be informed by feedback that he had provided in the past. If Bob is a completely new user then the system will rely only on the keywords for the initial set of results; this is called content-based filtering [24], [25]. Bob can select from that set or refine the results by providing more search terms. Bob may also reject the results altogether and completely quit the system. If an item is selected there will be a period during which Bob will take delivery and experience the item, shown as a dotted line between Phase 1 and Phase 2.

In Phase 2, the user provides feedback to the system regarding the item selected in Phase 1 to inform future recommendations. This is initiated with a prompt for ratings. Email is a common means for delivering a prompt. For movies, a prompt can come at the end of the presentation. Bob can refuse to provide the ratings and quit the system, but as this will end unconstructively with the data solicitation problem the desirable path requires that Bob elects to provide ratings. If Bob volunteers ratings, the system can then store the information and use it to match Bob to existing users in the community who share similar ratings. In turn, the learned preferences of those users inform future recommendations for Bob, shown by the dotted line at the bottom of Figure 1.

Collaborative filtering has been the most successful implementation of recommender systems due to the computational efficiency, as well as ***accuracy*** and ***diversity*** of the recommendations [2], [23]. Diversity is the extent to which the recommended set aligns with the preferences of the user, but the choices are different from one another [21]. Accuracy is measured by the difference between the ranking provided by the system and the ranking provided by the user for a particular choice, and is considered to be the principal measure of performance for recommender systems by developers [2]. If for example, in Figure 13, Bob's nearest community gives 4 stars to a particular television, and Bob gives the same television 2 stars after having accepted it as a recommendation, the accuracy would be $4 - 2$, or 2.

The Netflix Prize is a well-known example of the use of accuracy. The open competition offered $1,000,000 for a collaborative filtering algorithm that can improve the accuracy of recommendations for their movie streaming service [26]–[28]. Competitors were given a training set of 100,480,507 ratings from 480,189 users for 17,770 movies from the Netflix database. Another 1,408,789 ratings were concealed from the competitors as the test set, which represented users whose ratings the algorithms attempted to predict. The availability of the large initial data set made the problem tractable. Another advantage to the contestants was that explicit ratings were already made available to the contestants – it did not have to be collected from users.

In addition to explicit ratings, context aware recommender systems (CARS) incorporate contextual data furnished by the user to further refine the option sets. Content-based and collaborative filtering both serve as underlying processes in CARS; except contextual constraints are applied before the first set of options are presented to a user. Context pre-filtering require that the CARS be targeted to a specific application, so that the appropriate subject matter experts can be recruited to predefine the contextual categories, e.g., for self-driving car, weather, user preference, time of day, nearby points of interest, purpose of trip. Any contextual information that can be collected automatically, such as time of day, e.g., 4 pm Friday, is then applied to initial RS results before being presented to the user for additional contextual input, e.g., purpose of the trip is sightseeing. It is at this stage that the unique capabilities of the human, such as reading the intent of other actors, becomes a critical contribution to an operation. In the next chapter, I discuss the topic of CARS in further detail.

3. THE CONTEXT AWARE RECOMMENDER SYSTEM (CARS)

Context aware recommender systems (CARS) exploit human use of heuristics. It does this by incorporating contextual information that is selectively furnished by users based on their interpretation of the relevance of the information, e.g., limit options to quiet venues because work-related topics are likely to be discussed during the lunch, in addition to contextual information that is already accessible to the system, e.g., time of day. Context filtering is an iterative process requiring frequent engagement from the human to interpret and addresses real-time changes to the operational situation or environment, such as rerouting to pick-up a child from school triggered by the sudden unavailability of the original driver. When usefulness of the recommendations and overall interaction experience are taken into account, context aware RS are more effective than non-context aware RS [29]. Like collaborative filtering, the effectiveness of a CARS will be negatively impacted by data scarcity if an interface is unable to engage the user for data. Before the first set of recommendations are presented to a user, domain experts need to be involved in pre-defining the contextual data categories; the more of these constraints are added into the design of the RS, the more application specific the RS becomes. For example, a cinema CARS may have seating location as a constraint category.

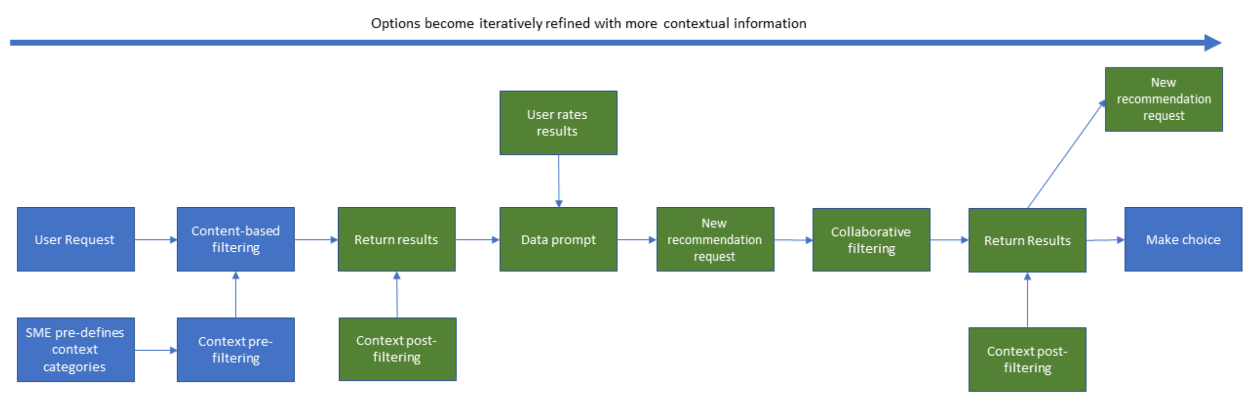


Figure 2. The contextual filtering process.

The context aware recommender system is composed of content filtering layers on top of the collaborative filtering process described in Figure 2. The process is again triggered by a request from the user which is subsequently passed through content pre-filtering to generate an initial set of options. The difference here is that a set of contextual constraints based on the categories previously defined by domain experts is applied to the results before presentation to the user. These initial constraints are typically attainable by the system and vary with time. Some examples are weather and geographical location; for transportation applications this information can be remaining fuel capacity. After receiving a request from the user and the context pre-filtering has been applied, an initial set of options is then presented to the user. The user will bring additional constraints, i.e., contextual post-filtering, to pare down the options, such as information about who and why they are making the request; returning to our self-driving car example, the information could be about real-time events that place new demands on the operation, e.g., a text message from home to pick-up more milk. In air transportation applications, a pilot may specify a medical emergency to limit options to candidate airports with medical facilities nearby [30]. After the contextual post-filtering, the system prompts the user to rate the options on-hand to generate a user profile for collaborative filtering in the next round of recommendation requests. The system then presents a new set of options on which the user can apply or remove contextual constraints in response to unfolding events that may impact the operation, or make a choice, ending the process. The strategy employed in CARS engages the user to improve the responsiveness of the recommender system to the operational environment. When user acceptability serves as success criteria for the RS, involving the user in the process by allowing them to manipulate inputs to view their impact on the solutions improves the likelihood of RS success. A key factor in encouraging user participation is the quality of the user interface experience.

The CARS process is an example of flexible rulemaking. Flexible rulemaking is the ability to add or remove constraints; I eluded to this in the examples given above. Conversational RS interfaces facilitate flexible rulemaking by allowing the user to manipulate RS input and preview the impact of the manipulation on the solution set [29]. This interaction strategy supports understanding of the solution space and in the process reveals reasons for the presented options. More importantly, conversational interfaces engage the human in an exchange of information with the RS as a necessary part of responding to real-time events in the operational environment. In off-nominal situations, this allows the user to identify which constraints need to be lifted to uncover options that would otherwise be inappropriate. For example, a pilot may need to override constraints that would eliminate highways or rivers in an emergency landing. In a similar scenario, when a failure requires an aircraft to land at the nearest airport, a pilot may want to insert a constraint to identify airports with suitable runway conditions given existing weather conditions [31]. Runway conditions and weather serve as contextual information that is not readily available to the system; the information must come from the pilot's experience with various airports or learned from colleagues. For recommender systems, ratings and contextual data are key human inputs. In the next chapter, I discuss these two sources of information further.

4. INPUT DATA ACQUISITION FOR RECOMMENDER SYSTEMS

4.1 Ratings Data and the Data Solicitation Problem

To achieve high accuracy, collaborative filtering depends on having many ***explicit ratings***, which are solicited from users by presenting them with Likert format queries for an item (e.g., how many stars do you give this movie?) [2], [21]. The benefits of collaborative filtering diminish when new items or users are introduced into the system for the first time, because the number of explicit ratings is not sufficient enough to create matches between the individual user and other users in the community [3], [4], [7], [22], [32]. If, for example, a user chose to exit the system after selecting an item instead of providing ratings when prompted, a profile would never be refined for this user, or created at all. The system would have to resort to alternative techniques, only to deliver sub-optimal recommendations.

Some existing solutions have been designed that deliberately expose the user to a poor initial set of recommendations and then depend on the chance that the user will continue to interact with the system to rate and refine the recommendations [33]. In most cases, the system never acquires the explicit ratings it needs to improve its accuracy because the poor recommendations diminish the system's utility and cause the users to abandon it early in the interaction loop [17], [18] – this is the ***data solicitation problem***. Also contributing to the likelihood that users abandon the system is the tendency for users to avoid providing explicit ratings due to the tediousness and unpleasantness of the task [8], [16]. Even for companies such as Netflix, who have a large database of explicit ratings from the community, this remains a problem because if individuals are new to the system and do not provide explicit ratings then they cannot be matched to other users in that community.

The use of ***implicit ratings*** has been employed as an alternative to explicit ratings because its collection is not imposing and does not require cognizant cooperation from users. Implicit rating sources come primarily from behavioral data such as the amount of time spent reading about an item or whether a user prints or saves an item [34]. These ratings have the disadvantage of requiring a greater volume of data to reduce inaccuracies in the inferences [2]. They are inaccurate because the behavior that is being observed is not always indicative of preference. For example, a record may show that a user dwelled on a webpage for a long period of time, but the user could have simply stepped away from the computer and neglected to close the web browser. Other sources of data have been mined to identify user preferences. These include contextual data from textual comments from blogs [4], or demographic information [35]. In the next section I describe contextual data in detail and discuss implications in vehicle operations for capturing human intent and purpose; a use case that demonstrates the relevancy of contextual data beyond user preferences.

4.2 Beyond Explicit Ratings: Contextual Information

Contextual information affords computing systems robustness across different settings[36]. For RSs embedded in mobile applications such as cell phones and cars, robustness means being able to produce appropriate solutions across different operational environments. Designers will need to consider two approaches to defining contextual information, positivist and phenomenological.

The positivist approach views contextual data as generally knowable and can be explicitly attained [36]. The data is representational and is considered peripheral to the task at hand. For example, the frequency of corrective inputs into a steering wheel would not be contextual data because it relates directly to a task; however, vehicle performance limits, geographical location, current lighting, and time, any information that can be explicitly given or received by the system [29], would be considered contextual data. The examples provided here, are fully observable. Partially observable data, on the other hand, contain some information that can be explicitly attained, but some parts that are latent. For example, a pedometer provides time and geographical location, but does not reveal if a user was walking on a treadmill or track, or if the reason for walking was for exercise or transiting between boutiques at a shopping mall. As implied by the examples above, the positivist approach to contextual information assumes that the information can be encoded and defined in advance. In addition, the positivist view assumes that the relevance of contextual data does not vary between activity or event, e.g., time captures duration regardless of it being for a football game or a commute to work. The positivist view dominates most computing research because of an existing bias towards data that can be readily applied to statistical trends and ideal mathematical models [36], e.g., IJtsma et al. [37], but this provides only a limited view of context. To better capture context, RS researchers have also adopted phenomenological perspectives. Users bring two categories of contextual information from the phenomenological view, latent and dynamic.

Latent contextual information consists of information that is discrete, but not observable [29], such as a recreational drive along a coastal highway; a machine in this example will not be able to detect that a route is intended for the enjoyment of scenery. The purpose of a trip will have to be explicitly conveyed by the human as part of the experience. From a phenomenological view, the route by itself does not serve as contextual information. If a human in the recreation drive example were to provide the contextual information, it would be in the form of preference information about various points of interest. With that preference information a route recommender can then provide different options based on trade-offs between a route that intersects fewer higher priority destinations, versus a route that intersects the most points of interest, but none of them being the most preferred. An example of the former would be a traveler expressing that he/she is interested in a wine tasting experience while on a road trip in the south of France; the recommender output would be a route that intersects the most wineries, ignoring length of trip. While for the latter, an example may be that the traveler decides instead that he/she is interested in site-seeing, in which case the recommender output would be a route that intersects items like national monuments, museums, architecture and some wineries as well. In both options, the route would be longer than options for a daily commute home, in which case the goal would be to minimize time and distance. Context in this example, is the relationship between the route and the nature of the experience – road trip versus commute. This relational property is one of four assumptions of the phenomenological view of context [36]. The second property is that contextual features are defined dynamically. Following on the previous example, a route may serve as context for a recreational drive in one instance, but in another instance that context can be replaced with certain buildings and architecture – you do not necessarily know in advance what people will consider context for a particular experience. The third property is that context is occasioned; a route that is context for a recreational drive on one occasion may be context for a daily commute in another occasion. The fourth property is that context emerges from activity; this property suggests that the route in the above example should always be associated with an experience to serve as context. This is a somewhat nuanced discrepancy with the positivist view that the route by itself is contextual information, and that the purpose for the use of the route is simply hidden or missing information.

The key takeaway from this chapter is that contextual information can be readily obtainable by a machine or, due to the relationship with experiences, it is attained only through engaging the user. The latter source of context motivates the interaction strategies for engaging RS interfaces – in particular, the implementation of conversational RSs. Given the ever-looming problem of data scarcity, RS developers will need to utilize any data available and accessible to them to generate effective recommendations. Next, we discuss methods for obtaining contextual data, followed by a description of concepts for joint action. Joint action naturally occurs when multiple actors interact for productivity or recreation; it motivates communication for exchange of information. An RS that can leverage joint action through an engaging interface can potentially be more effective at obtaining contextual information, as it emerges through the ongoing joint activity of generating options for shoes or routes in vehicles operations.

4.3 Obtaining Contextual Information

It is a key assumption in the present work that contextual information will have a positive impact on RS output. This assumption will need to be tested in future work. For now, one motivation for obtaining contextual data in real-time interaction is that it affords accommodating for changes in the operational environment that were not anticipated by the designer. Some of this contextual information is readily available to the machine, and some of this contextual information is latent to the machine. This hidden information must be extracted from ill-defined models of how users understand the ways a system works and how it interacts with the environment. These internal models are difficult to define because they vary by the unique experiences individuals have for an activity such as driving [38]. Humans can infer context information from each other if they share common mental models; if there is no common mental model, then people simply ask each other or the information surfaces during ongoing dialogue. A machine can employ similar means to obtain contextual information; Adomavicius et al. [29] provides three ways: explicit, implicit, and through inference.

Contextual information that is obtained explicitly is collected by asking the user directly [29]. For website applications this can be done with a questionnaire that is completed by a user. Alternatively, the data can be collected from readily available sources like time and location from GPS. Contextual data that is obtained implicitly is an extension on data collected from available sources, such as identifying that a person is in a national park based on GPS information. When contextual data is collected by inference, statistical or data mining methods are employed. A prime example of this is the collaborative filtering technique described in Chapter 2, where the preference for an item can be predicted based on ratings given to items purchased in the past. Inference techniques can also be applied to textual data, e.g., social media dialogue, to identify common symptoms reported by users that might be related to an increase in hospital visits within a city.

Explicit and inferential means of obtaining contextual data require direct input, at least initially, from the human, which makes them susceptible to the data solicitation problem. This problem is not solely a burden on the machine. A considerable number of factors impact whether humans will divulge information to one another, such as trust and the level of risk that the requested information carries [39]. However, those factors considered, active dialogue creates opportunities to engage and build the trust required for information exchange [40]. From a phenomenological view, contextual information emerges when people engage in activities together; these activities include both recreational and productivity tasks for work related roles where communication is required. Significant advances in automation will need to happen before machines can leverage engagement for obtaining information from humans. Factors that influence human-human engagement, such as trust, will also influence continued engagement with a machine, e.g., [24], [41], [42]. I believe these advances will come in two forms: the ability for machines to engage in joint activity with the human, as well as mimic human verbal and visual communication. I begin the next chapter with an introduction to joint activity, followed by a discussion about engagement where I define the construct. I then begin a discussion that will continue through Chapter 6 about human computer interaction concepts that can potentially influence engagement and joint action, which includes mimicking human communication.

5. HUMAN COMPUTER INTERACTION APPROACHES TO ENGAGEMENT

In this chapter, I begin with defining joint activity in Section 5.1. In general, joint activity is any activity involving more than one individual and will always involve social elements where communication and coordination is taking place. The concept of joint activity is important to this work because the exchange of information occurring within joint activity creates opportunities to solicit data from users while interacting with recommender systems. Engagement can feed into the joint activity loop by encouraging actors to continue interacting with one another. For the topic of joint activity, I will define the concept, describe two different types of joint activity and the conditions that determine when users will pursue joint activity. Next, I will take a deeper dive on engagement, starting with a working definition of the construct in Section 5.2. I will cover various frameworks for engagement from the literature, so that I can point to the factors that determine engagement. This will help to focus what interface implementation approaches should be studied for their potential impact on engagement. These approaches will be covered in Section 5.4. Finally, the concept of trust is important to this work because of its potential influence on whether users will interact with machines for joint activity and whether engagement will result from the interaction. I cover trust in Section 5.3 where I define the construct and discuss its relationship to how users will interact with machines, as well as discuss how system transparency serves as a determinant of trust.

5.1 Joint Activity

The concept of joint activity is important to this thesis because it drives the communication needed to acquire information from people. Clark [43] wrote extensively on joint activity, defining it to be an activity done in coordination by two or more actors. Joint activity can be recreational or practical. Recreational joint activity can be gossiping between two or more people, where the outcome is social reward. Practical activity is when two or more people interact to accomplish a task, such as editing a book. Clark added that in a joint activity the actors can have roles that determine how labor would be divided, and that the actors can have private goals, as well as shared goals. Clark also conveyed that joint activity is composed of nested joint actions. Finally, while joint activities have a start and end time, they can happen simultaneously, or intermittently, as when a family watches television while having a meal together, and then periodically engage in conversation as different topics in the news are reported. Clark believed that it is impossible to have joint action without communication and vice versa. When actors engage in communication, they bring context, as in prior knowledge, beliefs, and assumptions.

Johnson et al. [44] proposed that interdependent relationships are formed for joint activity within a given context. The relationships are driven by a need for actors to make-up for gaps in individual knowledge and skills to perform an activity within a context. Performing actions requires skills and knowledge. Johnson referred to these individual skills and knowledge as capacity, and when an actor lacked capacity to perform an activity within a given context, they were considered dependent. Another reason for forming interdependent relationships is to take advantage of the combined capacity of independent actors to enhance performance of an activity. Independent actors carry all the capacity needed to accomplish a task - within a given context; that is, given other contexts, the independent actor can lack capacity and be a dependent actor. A surgeon can exercise independence when diagnosing a patient but will require a team of technicians to perform surgery. Context is important. Johnson et al. defined interdependent relationships based on functional need. However, it is also important to note that interdependence can also be non-functional, such as the need to seek the consolation of others upon the passing of a loved one, and the desire for others to express sympathy. Consistent with the phenomenological view, Johnson cited [45] that the relationships formed for joint action shape context that is internal to the human, such as knowledge, beliefs, and assumptions. For functional interdependence, that could be knowledge about how to operate a word processor, and for non-functional interdependence the knowledge can be how to convey sympathy that will be acceptable to another individual.

Based on Johnson et al., [44] an RS implementation that would be effective at engaging users for obtaining information would then need to provide capabilities that users need to accomplish an activity, or generate options that effectively leverage the context information provided through joint action with the human. RS designers have generally focused on fostering practical engagement by demonstrating such utility to users so that they do not abandon the system before it has obtained the information it needs to continue serving useful recommendations. From a utility perspective, a user's likelihood of continued engagement depends on early experience with how well a system performs its intended function [40], [46].

In addition to these practical reasons, Clark [43] suggested that joint activity includes verbal interaction, like attending a lecture or gossiping, to fulfill less tangible goals such as intellectual or social reward. To achieve this level of social engagement is very difficult with machines. It requires mimicking human-to-human verbal interaction; the likes that are only recently being explored through advancements in artificial intelligence. Thus, before effort is applied to developing machines capable of social discourse, the principle assumption that social discourse will lead to more contextual information from users will need to be tested. In advance of evidence that this assumption is true, however, there is evidence that users have an inherent tendency to socialize technology. In the next section, I describe how this already present tendency to socialize technology can be leveraged to solicit social engagement. Thereafter, I return to the topic of practical engagement with some background on early human and machine interaction strategies, and then present more recent developments in interaction design that leverage teaming principles that center around joint activity.

5.2 What is Engagement?

There is no widely accepted definition of engagement, however, the work of O'Brien and Tom [14] provides a definition that is helpful for discussing recommender systems. From a systematic review of the literature and semi-structured interviews with users, O'Brien and Tom identified attributes commonly used to describe engagement. While engaged these attributes are expected to impact a user's experience at various intensities. They proposed that **engagement is a user experience that can be described as challenging, positive in affect, endurable, aesthetic and appealing to the senses, appropriate and responsive in feedback, novel, interactive, provides a sense of control, interesting, provides a sense of awareness, and meets user motivations**. Engaged users report they are **challenged** when the application compels them to invest resources and effort into the outcome [47]; for example, designers of computer games will put puzzles and tasks on a user's critical path to gaining levels and items that will afford them an advantage over their opponents. Recommender systems can challenge users to provide feedback and ratings for products by offering discounts or purchase credits in exchange. **Positive affect** is any positive emotions that results from interacting with the technology, such as happiness, satisfaction, or a sense of accomplishment. As a nuance, positive affect can result from engagement that was retained from negative affect. A challenging computer game, for example, can result in user frustration; however, if the challenge is appropriately calibrated so that it is possible for the user to emerge from the frustration and succeed, the overall experience can conclude with positive affect. Interactions with recommender systems can result in positive affect in the form of satisfaction if the system generates products and services that meet a user's current needs. User experiences are **endurable** if they report that they would revisit the technology, recalling that an experience with an application was positive. **Aesthetic and sensory appeal** is the user response to more static elements of technology, such as visual or auditory presentation [48]. In both usability and design fields, engagement comes from the additive qualities of a user interface that make it pleasant and satisfying to use, such as the colors and spatial design that set tone and style. Visual appeal can come from a computer agent that is presented with attractive human features. Auditory appeal can come from a desirable soundtrack for a computer game. Aesthetics, how people respond emotionally to the pleasantness or beauty of things including objects, people, events, or recommender systems, influence our willingness to interact with them. Referring to their product design aesthetics, the Swedish sports equipment company, POC, claims that "…the best or the safest helmet is one that somebody chooses to wear" [49]. This example attempts to show that aesthetics can be applied to technology to influence users to exercise functional behaviors – in this case it is to wear head protection. For recommender systems that involve verbal interactions, such as Amazon Alexa, aesthetic and sensory appeal can be a response to the tone, language, or accent (e.g., British) implemented for responding to commands from the user. The quality of the **feedback** to the user can positively influence engagement [50]. Feedback to the user supports engagement when it is perceived as prompt, accurate, and appropriate to the context. Like communications between humans, an interaction with technology that possesses good feedback responds to user input with a minimal amount of delay. For example, a good system error notification should come immediately after the action that caused the error and should clearly indicate what error had occurred, what system features have been impacted and what actions caused it. Feedback that supports engagement is not the cryptic and canned, such as responses that often come with computer error messages; "sorry we are not able to process your request right now; please try again later". Unfortunately, many recommender systems that employ verbal interaction designs, return error messages are that are equally cryptic, “I’m having trouble understanding right now.” **Variety and novelty** reflect engagement when the users believe variations in their inputs lead to different outcomes and interactions not previously experienced. In computer games, this can mean unexpected visual and auditory events, such as encountering an enemy as a player turns a corner in a first-person shooting game. With recommender systems, this means discovering new products and services. Diversifying the actions one can take with the interface, as well as creating interactions that change with context [51] enables exploration and facilitates variety and novelty. Google and Alexa assistant devices demonstrate this by supporting recreational and practical interactions, such as delivering music with voice commands, reporting the weather, activating household appliances, and then by proactively interacting with the user, such as requesting that a command be repeated if the initial input from the user is unclear or ambiguous, or an RS informing a user of new buying opportunities that they had not previously considered. Users find **interactivity** if the technology supports active exchanges with the user. Movies are not considered interactive. Although one may be perceptually and cognitively active while watching a movie, presumably, there is no direct relationship between a user’s cognitive state or actions and what is presented as the film narrative unfolds. Video games are the most common example of interactivity, and test drives for a new car can be another example. Voice recommender systems support interactivity when they offer further assistance that builds on a previous command. When asking Alexa to play a song it may respond with, “Music is streaming on another device, would you like me to play music here instead?” **Perceived control** is achieved when users believe they are influencing outcomes through interactions with the system. As demonstrated with Google and Alexa assistants, providing a greater range of control to the user can have a positive impact on engagement, for example, allowing the user to order products by voice or query the status of an order, “Alexa, where is my stuff?” In games, engagement from perceived control can come from allowing the user to exert more effort to attain reward [51]. For recommender systems, a designer can influence a user’s decisions by manipulating what items are presented in the list of options, but the user is able to choose among those options. It is instructive to note that actual control and perceived control can be dramatically different. If the options are rigged to direct a user to a particular solution, but the user is not aware of this, the user may still have the perception of control. If a recommender system provides an explanation for the set of presented options that relates back to a user’s data input, this can have a greater impact on perceived control; here, the user is being informed that by way of their input data, they are directly influencing the solution set [52], [53]. Deceptive manipulations of perceived control risk being discovered and losing the user's trust (trust is discussed in section 5.3). A user is displaying **interest** when they are willing to attend to particular topics or content with an application [54] . If a user's interests relate to cooking, and an application delivers content that is relevant, such as recipe videos, a user may be expressing interest by browsing through the videos. Similarly, users may display interest in a recommender system for route navigation by interacting with system filters to reveal various points of interest during a road trip. Engaged users remain **aware** of other people and events in the surrounding physical world [14]. Due to this awareness, it is likely that a user's interaction with a system will be periodically interrupted, but the users can be resistant to this interruption and continue to interact with a system shortly after. Thus, it is reasonable to suggest that one indication of the intensity of engagement is a user's sustained interaction with an application, despite intermittent interruptions. As implied by the endurability attribute defined above, an application can be so engaging that a user may seek to revisit the interaction even after a long hiatus. Finally, a system interaction that results in engagement satisfies user **motivation** by meeting user's intents and purposes for entering interaction. These intents and purposes can be productive, like creating a document, or recreational, such as drawing and painting[1]. I reserve a more detailed discussion of the construct of motivation for later in this section.

Based on the above attributes, O'Brien and Tom [14] proposed a model that describes engagement as a process that begins with an entry point for engagement and then transitions to engagement itself and ends with disengagement (Figure 3). Importantly, the process can then be reinitiated with a user revisiting the system to re-engage in the near or distant future, but this is not guaranteed. At the point of engagement, a user may begin interacting with a system to determine if further interactions are worthwhile. At this phase, the attributes that may lead to transitioning into engagement are aesthetics, novelty, interest, motivation, and whether the user believes the interaction will address practical or experiential goals. Many of the interface features that will influence these attributes will be passive. For example, the artistic design of the technology may

[1] Identifying whether an activity is practical or recreational would, in most cases, be a difficult task if context is not considered. For example, drawing and painting if done as a profession would be a practical activity. However, creating a document to serve as a diary or to post as a journal of someone's vacation travel can be considered recreational.

have a positive impact on the user's aesthetic response. Another example would be if the interface reflects novelty by conveying a rich interactive experience through multiple options and features and different modes of interaction., e.g., visual and tactile. Also, at the point of engagement, features of the interface that align with the user's interest or current motivation, e.g., a functional motivation to accomplish work, can increase the likelihood that a user will become fully engaged in the technology. Finally, addressing practical goals means that a recommender system directly conveys that the main purpose of the system is to meet a need like identifying a physician that is within a user's health plan network, or if a user's goal is experiential, the recommender system may employ a human-like virtual agent to make an interaction appear more natural.

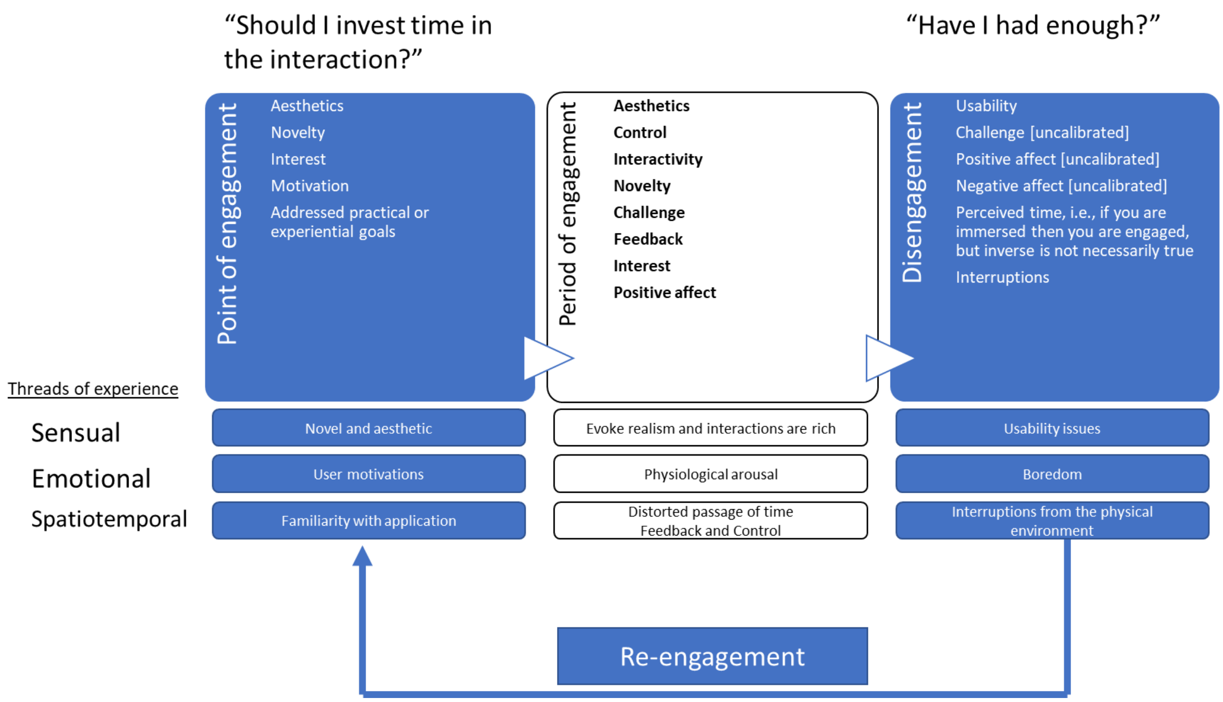


Figure 3. The O'Brien and Tom [14] process model for user engagement.

The end of the engagement process is composed of attributes that contribute to disengagement. Users will disengage if the technology's interface is not usable; e.g., features are difficult to find and if there is too much response delay to user actions. Users will also disengage if the challenge presented by the interaction is uncalibrated [47]; that is, for example, designers of a computer game are not sensitive to users' skill levels and make a task too difficult to achieve. Both uncalibrated positive and negative affect can increase the likelihood of someone disengaging a system [14]. When positive affect is uncalibrated a user will quit a system if they satisfy motivations too early in the interaction. In contrast, when negative affect is uncalibrated, an experience is so challenging that a user quits prematurely due to frustration.

Perceived time is somewhat nuanced here. A complete discussion of time perception is out of scope, but it is worth addressing briefly because it provides some empirical grounding for what users are experiencing during engagement. From that body of work, Vierordt's Law [55] can be applied to understand time distortions occurring with engagement. According to Vierordt's Law, an observer will tend to overestimate the elapsed time for short intervals of time and underestimate the elapsed time with long intervals. The distortion in time in engagement [56] can be an indication that a user may continue to interact with an application for longer durations, thus delaying disengagement. While engaged, a user is likely to allow a significant amount of time to elapse before checking the time; in this situation, Vierordt's Law predicts that in retrospect the user will likely judge that the time elapsed has gone by quickly. Alternatively, if the interaction is not engaging, like in the case of boredom, the user may check the time more frequently and judge that they have invested too much time with an interaction and quit the system. Thus, distortions in perceived time can occur when a user is engaged or not engaged. The amount of time a user allows to elapse before checking the time can be used to determine how valuable the interaction is to the user in term time resources. Unlike with immersion, users maintain an awareness of the physical world while engaged to constantly evaluate whether they should remain engaged or if they should disengage. Interruptions from external events such as needing to eat or sleep, meeting a scheduled dinner date, or stopping to satisfy the attentional needs of a toddler create opportunities to assess the value of continued interaction with a system. While fully immersed, many of the aforementioned events may be ignored or unnoticed. When immersed a user has committed all their resources into the interaction, and in the case of addiction, the user has committed more then he/she is capable of committing.

Re-engagement carries the same attributes identified at the point of engagement. However, experience with the system plays a large role. If designers hope to re-engage users, they need to entice those users to revisit the system with the promise of another positive experience. The time between disengagement and re-engagement can be long in duration, where a user may leave the system for days or months before returning to it. Alternatively, there could be multiple instances of disengagement and re-engagement within the same sitting. O'Brien and Tom [14] refer to this as engagement episodes within a single session, and assert that this pattern of re-engagement is intrinsic to their model. This likelihood of returning to interact with the system can serve as a strong indication of engagement and subsequently serve as criteria for a successful system [57]. For recommender systems increasing the user's exposure to the system creates opportunities for user data acquisition.

O'Brien and Tom [14] proposed that engagement represents only a part of the overall experience with technology; other threads of experience can also be present. The attributes within the threads change with the different phases of engagement: point of engagement; period of engagement; and disengagement. Three were summarized in their model: sensual; emotional; and spatiotemporal. I believe it is important to discuss the threads here because they can serve as additional indicators of transitions between engagement phases. For example, attributes of the emotional thread include physiological arousal during the period of engagement. Presuming this relationship exists, researchers can apply heart rate as an index of engagement. I describe each of these threads of experience next.

The sensual thread of experience relates to the visual, auditory, and interactive components of the user's interaction with technology. At the point of engagement, the user response attributes associated with those components are novelty and aesthetics. During the period of engagement, the sensual experience is characterized by a sense of realism, sustained interest, and the interactions are perceived as rich. As this is the sensual thread of experience the perceived realism and interest will come from graphical, auditory, tactile or olfactory elements of the application. The richness in interaction facilitating the sensual experience, as O'Brien and Tom [14]described it, will depend on the customizability of the interface. Users will disengage the sensual experience if the usability of the application is poor. That is, it is too difficult to interact with the features of the user interface and the user is inhibited access to features of the application that provide the sensory experiences. Alternatively, users can elect to disengage if their motivations are satisfied, such as playing a role-playing game until its conclusion, or they can quit a game prematurely because the level of difficulty arouses too much frustration. Sensory experiences can also vary in quality, and depending on the activity, the absence of a sensory experience can lead to termination of the interaction altogether. For example, some driving simulators provide haptic feedback to steering controls to augment the visual and auditory experience. A user can continue to interact with the simulator with a diminished sensual experience if the haptics were to fail, or if there is an accurate pairing of the haptic feedback with the conditions of the road. However, if the visual elements of the interaction were lost, such a frozen or blacked out view of the road, the user may elect to completely abandon the interaction, thus terminating the sensual experience. With the emotional thread of experience, users experience positive emotions, such as enjoyment, satisfaction, and fun. During the period of engagement user motivations are the major attribute for continued interaction with the system. As described previously, user motivations can include a desire to feel accomplished. If the application provides a feeling of productivity, then this functional motivation may be satisfied. During the period of engagement users' emotional experiences may include feelings of enjoyment and fun, as well as physiological arousal, e.g., increased heart rate and pupil dilation. At the stage of disengagement, the emotional experience is lost when people have satisfied their motivations or have exhausted the novelty of the features and have become bored. As suggested by the term, the spatiotemporal thread of experience is related to the time and space of the interaction. This is characterized by people's perception of time, awareness of their own internal states and the environment. From the interviews conducted by O'Brien and Tom [14] users reported that the spatiotemporal aspects of their experience were marked by a distorted sense of time. For the period of engagement, that distortion was that the time they spent with application appeared to have gone by quickly. Unlike with immersion, the time distortion here is not accompanied by a complete lack of awareness of the physical world. A user may still be keeping track of time when engaged. In addition, a distorted sense of time, users have diminished awareness of their own internal states; this includes hunger or fatigue. Users also reported that they will have a diminished sense of their surroundings, such as not noticing that it has gotten dark outside. At the point of engagement, the spatiotemporal experience attribute is the sense of familiarity or acclimation to the interface that is gained from the initial time spent exploring an unfamiliar interface or a new feature. During the period of engagement, the attributes include a distorted passage of time as mentioned above, a sense that the user is receiving appropriate feedback and control from the interaction and then a lack of awareness of the physical surroundings, but a strong awareness of other people if the user is interacting with other users within the application[2]. Finally, the spatiotemporal experience will conclude if there is a compelling interruption occurring in the physical environment - perhaps a toddler requesting the attention of the engaged user.

The engagement model proposed by O'Brien and Tom [14] is a convenient starting point for a discussion about defining engagement. It assumes that engagement goes through different stages. Each of these stages are characterized by different or overlapping attributes that reflect users' responses to the interaction. Although a definition of engagement is relevant to reaching a common objective for the design of applications such as recommender systems, a definition alone does not provide much guidance for implementation. For that, we turn to the work by Perski et al. [13] and Short et al. [58] who proposed frameworks that identify determinants of engagement. Although the terminology between their frameworks vary, the constructs they introduced are similar. I discuss them together next.

Perski et al. [13] proposed a definition of engagement that distills the list to just three commonly used attributes: attention; interest; and affect while proposing to operationalize engagement to reflect the behavioral aspects of engagement. This operationalization is grounded in the physical manifestations of the engagement experience. In manual driving, engagement has been treated as a binary event occurring when a human operator is applying cognitive and manual effort to operate a vehicle [59]; according to some researchers, when that is not happening, the driver is considered disengaged. However, the Perski et al. [13] definition for engagement supports both operational and experiential indicators of engagement. The definition is as follows, "Engagement with [Digital Behavioral Change Interventions] is (1) the extent of usage (e.g., amount, frequency, duration, depth of usage) and (2) subjective experience characterized by attention, interest and affect." Digital Behavior Change Interventions (DBCI) is defined as a product or service that uses computer technology to influence human behavior [60]; for example, a software application used to set reminders for taking medication.

[2] Although out of scope, presence [54] is the construct that describes the phenomenon of having a strong awareness of other people while interacting with them over an application as a medium. Presence can be generated by immersive technology, but it is treated in literature as a separate construct. Here presence is treated as an attribute of engagement.

The discrepancy in the number of attributes describing engagement between O'Brien and Tom [14] and Perski et al. [13] comes from the number of application areas considered in their definition of engagement. In O'Brien and Tom, the goal was a more general use of the term engagement, so their definition was drawn from studies which focused on a variety of applications like video games, educational applications, online shopping, and web searching. These applications may have been designed to help the user achieve specific goals for enjoyment or accomplishing a task. Perski et al. [13] and Short et al. [58] were interested in defining engagement for DBCI, where the design of such interactions carried the explicit goal of influencing the behavior of the user, such as quitting smoking; in the case of the current work, this means volunteering data for improving recommender systems. An engaging interaction would be the motivation for achieving those design goals. As implied in the definition proposed by Perski et al., an engaging interface extends the usage of a system. For recommender systems, enticing users to increase their exposure to a system affords more opportunities to acquire user data. In addition to identifying the attributes of engagement, Perski et al. and Short et al. identified direct and indirect determinants of engagement. These determinants are relevant to this work because they provide guidance for how to tangibly solicit engagement through interface design.

In the model proposed by Perski et al. [13], there are two main determinants of engagement[3] (Figure 4). The first comes from the interface implementation concepts that are used to create the DBCI; this is of most interest to the current work because of the relevance to RS interface design. The interface implementation concepts that comprise the DBCI interface are categorized as delivery or content. The second direct determinant of engagement is context. Perski defines this as the product of population attributes and setting attributes. The population attributes can be divided into demographic and psychological constraints associated with the users. The setting attributes can be divided into physical and social constraints. Perski et al. listed many attributes for each of these attributes, but I focus only on attributes relevant to this work.

[3] Perski et al. [13] included constructs with had evidence-based relationships with engagement, as well as constructs with a hypothesized relationship to engagement. In this document I present only the former.

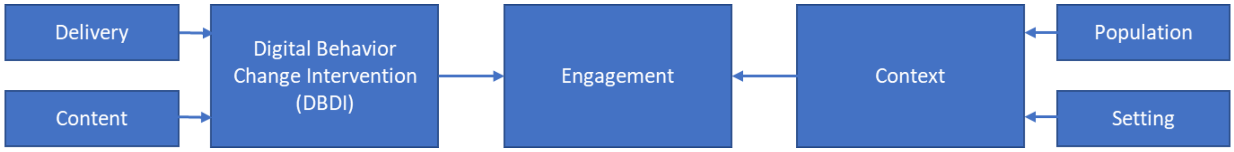


Figure 4. Determinants of engagement: a simplified version of the Perski et al. [13] framework.

For purposes of understanding engagement, the DCBI **delivery** attribute is defined by message tone and narrative. Message tone is the terminology and wording used to communicate with the user. Narrative is the presence of a storyline. Both of these have been found to positively influence engagement and have implications for designing computer agents that can appropriately communicate with users and that can mimic human social discourse by sharing a backstory about itself, even if the backstory is fictitious. For human-machine interfaces that include an embodied human-like agent, telling stories that reduce social distance with listeners can improve engagement [61]. Likewise for DBCI, the **content** attribute is defined by concrete content such as goal setting and reminders, which are designed to help users achieve goals like losing weight [13], [58]. For recommender systems, content would be route options for navigation applications or movie options for streaming applications. In addition to the foregoing, Short et al. [58] included the usability of the system, (the quality of an interface that is associated with ease of use, reliability of the functions of an application, and how quickly and easy it is to learn how to use an interface [62], and personal relevance of the functions and content of the application, as part of what they called interventions, i.e., DBCI, that determine engagement; both of these are defined in the model by O'Brien and Tom [14]. Usability would fit into the delivery aspect of the framework by Perski et al. [13] and personal relevance would likely fall under content. Next, I discuss context as a determinant of engagement in the model proposed by Perski et al.

Prerequisites to engagement are not one-size-fits-all. Different activities elicit engagement for different populations of people. Thus, what Perski et al. [13] refer to as **population** attributes are common attributes of individuals that influence the likelihood of engagement with an application. The population attribute is related to human characteristics like demographics and psychological states. For demographics, these are age, gender, ethnicity, etc. For psychological constraints, these are motivation, need for cognition, and interest; these were defined in the model by O'Brien and Tom [14]. Seah and Cairns [63] share a similar perspective on population as a determinant of engagement. For Seah and Cairns, engagement is a stable characteristic of an individual, where an engaging person is someone approachable and capable of capturing and holding interaction with others for long durations. Thus, it appears engagement is more likely with individuals who possess characteristics that expose them to opportunities to be engaged.

The **setting** attribute includes physical constraints, such as the amount of interruptions occurring during interactions with an application, time available to interact with the application, and access, e.g., internet connection for an online web application. Setting also includes social constraints. These constraints include culture, norms, and social cues. For a virtual travel agent, all these constraints will likely apply if the agent is expected to successfully solicit engagement. For example, would prolong eye contact be culturally inappropriate to some users? For norms, there may be common practices for pointing or gesturing to information on a display. Finally, a virtual agent may provide a social cue for sustained engagement with eye contact and then cue disengagement by looking away or down.

The model proposed by O'Brien and Tom [14] and the framework from Perski et al. [13] and Short et al. [58] provide a starting point for defining engagement. The attributes they have identified can serve as design objectives for applications that aim to influence user engagement. For example, capturing user interest as an attribute of engagement can serve to justify personalization of content for recommender systems, as opposed to adhering to simpler content-based filtering techniques. By identifying the determinants of engagement designers can then begin specifying what features and capabilities are required for achieving user engagement. This can involve modifying a virtual agent to employ small talk when users are from cultures that prefer more social interaction versus functional interactions. Next, I elaborate on why engagement is important, and then I narrow my focus on social determinants of engagement because of its relevance to communication with users for soliciting data from them.

Doherty and Doherty [62] in an extensive literature review, proposed that engagement is important for three reasons: 1) conducting basic research to understand engagement as a construct; 2) designing products to elicit engagement; and 3) designing systems that elicit engagement and then leveraging that engagement to achieve other goals. Reason three is the motivation for the present work and was also the intention of the definition proposed by Perski et al. [13]. Engagement has been identified as a prerequisite to achieving a variety of system objectives [15], such as improving perceived product utility and obtaining healthcare information, because engaging experiences can motivate continued user interaction with those systems. It can also serve as an intermediary to achieving learning, encouraging healthy behavior and well-being for individuals. For healthy behavior, an application can make recommendations for a healthier selection of food. If the interaction with the application is engaging, users may be more likely to take the system recommendations and change their diet. Provided valid indexes, engagement can potentially direct interface design by identifying what elements improve user experience [14]. In the case of recommender filtering, the primary application area of interest here, engagement may be the prerequisite for acquiring ratings and context information from individuals. Engagement can serve as part of a feedback loop to encourage continued interaction with a system, where an interface that successfully elicits engagement entices users to return to it [13]. I reason that the longer users interact with the system the more opportunities are afforded to solicit and obtain information. As mentioned earlier in this section, one of the key determinants for initiating sustaining interaction with a system is motivation. I discuss motivation in detail next.

The experience of engagement can result from the satisfaction of social, hedonic, or functional motivations [51]. Social motivation is the desire to connect and share with other people, such as the type of engagement people experience on social media, where recommendations to connect with other users can increase the likelihood of someone revisiting a system such as Facebook to post more personal pictures or comments [64]. Hedonic motivation is a desire for activities to be fun, enjoyable, and pleasurable, which, in the research, is typically associated with gaming, but can also be extended to learning and education for school children [12], [65], or dining at a gourmet restaurant. These activities can be goal-directed, like in computer games, or not goal-directed, such as enjoying a symphony orchestra or a play. A functional motivation is a desire for ease of use with technology, based on criteria such as efficiency (e.g., fewer mouse clicks to navigate to a webpage), ease of use, and time savings from use [66], as opposed to not using the technology. Similarly, functional considerations can be made to interactions with other people, based on ease of communication and personal compatibility in terms of character and values. In functional activities, the goals and objectives are explicit.

As an outcome of interaction with the physical or virtual environment, engagement is an individual experience that can come from two types of antecedent activity – practical or recreational (Table 1). These activities can be motivated by social, hedonic and functional motivations. Intuitively, one may expect engagement to emerge with an appropriate alignment of activity to motivation. For example, an individual who is hedonistically motivated would seek recreational activity, while an individual who is functionally motivated may seek practical activities. With social motivations the relationship with activity can be less distinct, as this can be satisfied with recreational or practical activity. The relationship between activity and motivation has implications for RS interface design. I provide further treatment of this topic below to discuss the nuances in the relationship between activity and motivation.

Table 1. Engagement emerges from an appropriate alignment of activity to motivations.

		Motivation		
		Functional	Hedonic	Social
Activity	Recreational		Engagement	Engagement
	Practical	Engagement		Engagement

Practical and recreational activity was defined in Section 5.1; I review them here within the context of user motivation. Practical activities are goal-oriented, and aimed at productivity [43]. Practical activity includes collaboration between people on a book, building construction, or developing software applications. This kind of activity can result in satisfaction of a functional motivation (e.g., perceived competence of other people), but social motivations can occur (e.g., desire for one's opinions to be validated), as well, when the task-oriented activity requires a considerable amount of coordination and communication. In interactions between people, practical activity and social discourse rarely, if at all, can occur in isolation. When engagement results from interacting with technology, the experience can reflect ease of use of the interface, but socialization has been evident to varying degrees with technology, as well [11]. When socialization has occurred, users treat the technology as if it was another human actor. The most familiar example of socialization of technology is when people express anger and frustration to a computer when it fails to function as expected. Recreational activities are activities for enjoyment with both other people and technology [43]. Outdoor sports and online gaming are examples of this. As stated earlier, hedonic motivations can be satisfied with recreational activity. Social motivation can also contribute to recreational engagement from activities like gossip or a political debate, where there is a desire to appear credible to others. In gaming the social motivation can come from the desire to rank above peers in performance. When the games are being played as a team, such as football and basketball, there could be a desire to seek praise or negative attention by insulting others who have been defeated. Similarly, social motivation can be satisfied with practical activity, as with a team of graduate students conducting a study or a team of construction workers assembling the foundation for a building. Unlike with functional and hedonic motivations, where the quality of the technology interface may be the most responsible for the quality of the engagement that results from the interaction, the potential engagement that can be achieved when the interaction is socially motivated also depends on the quality of the interaction with other actors, as well as the technological medium. A determinant of the quality of social interaction with other actors are the social discourse practices that are employed. For humans, barring any developmental deficiencies, e.g., autism, these social rules are learned with continued interactions with each other. If a machine can be designed to employ social discourse practices, it may also leverage some of the advantages of engagement as a result (the topic of social discourse will be discussed in Section 5.4.2).

In this work, the principle claim is that by making an interface engaging, one can improve the likelihood of obtaining data from users. This section was devoted to defining engagement and identifying how a designer can approach developing interfaces to elicit engagement. These approaches address both functional aspects of an interface, such as ease of use, and interaction schemes that mimic human social discourse. In the remainder of this document I focus on discussing concepts that support the latter interaction scheme, mimicking human social discourse, because of the potential for soliciting engagement and subsequently improving the likelihood of obtain data from users.

5.3 Creating Narratives

Mallon and Webb [67] argued that narratives play an important role in engagement. Narratives in gameplay are stories in multi-objective role-playing games where the player interacts with virtual worlds and the virtual agents within them from a first- or third-person perspective. Narratives are important to this work because they have an impact on engagement by stimulating important attributes associated with engagement. These attributes were discussed in the previous section and they are interactivity, control, and positive affect. From a focus group study, Mallon and Webb derived six different interaction design practices, called propositions in their work, to make-up a narrative. They referred to these practices as **spatial containment, causality, skill-based interaction, causality of dialogue, illusion of intelligence** and **invisibility of the medium**. I review each of these propositions next.

Spatial containment refers to limiting the discovery space so that it is not perceived as too large to the user; in virtual worlds this is done by constraining the setting of the narrative to rooms, islands or scenes. However, this does not necessarily mean that the depth of the interaction of the elements within that world must be limited. Researchers have advocated for maximizing what they call is the richness of the interaction [54], which is a very difficult technical undertaking. For the purpose of eliciting engagement, any action taken with an object or agent should be diverse and result in different action paths based on the context of the narrative. For example, when speaking to a machine agent about "grabbing a bite" during some time in the evening, the machine may respond with, "Sure, I've got a couple of places in mind that have some great reviews for dinner. Should we dine-in or take-out?" The machine on another occasion may instead say, "No problem, should we go to your usual for dinner, or do you want to try something new?" For richness in interaction, the machine should be able to access a very large variety of topics in response to the human in a conversation. Spatial containment may solicit engagement by calibrating the challenge [14] associated with the activity, so that success with activity is within the capabilities of the user to achieve, resulting in a more positive experience for the user.

The second proposition is to **create causal connections** in the narrative. For Mallon and Web [67], this meant providing clues to solve tasks, such that they are woven into the narrative. In games, this could mean that you click on different objects in a scene to reveal parts of the story and then clicking on a character within the game would prompt it to provide context for the object and the story within it. The information can then be used to determine if the object can be used to open a box or a room from a different scene, but the users cannot be explicitly told this. They would be expected to conclude it themselves. Creating causal connections can solicit engagement by satisfying interactivity and novelty requirements [14] of the construct. That is, it prevents receptive responses from the system that would otherwise make the experience boring.

The third proposition is **skill-based interaction** [67], where the user exhibits motor skills to achieve system objectives. This proposition was originally applied to the motor-skills exhibited in a video game, like target acquisition and character animation control. An extension of this would be to elicit motor action for more engagement. An example would be augmented reality systems such as Microsoft Hololens that provide users the ability to manipulate virtual objects to acquire information about real-world landmarks. In initiating dialogue with a machine, this can mean waving to enable the microphone. These do not require considerable skill but could potentially magnify engagement by involving the user's motor abilities. Skill-based interactions can positively impact engagement by satisfying the control attribute [14] of engagement.

The fourth proposition is causality of **dialogue** [67]. This means that the system should employ dialogue that is goal directed and assists with action, or joint activity. The dialogue can be a backstory or humor. For a conversational movie recommender system, the backstory can be about a fictitious experience with a certain movie genre that leads up to a request for information about what movies the user prefers. With dialogue a system meets the interactivity attribute of engagement.

Proposition five is the **illusion of intelligence** [67]. This proposition carries the most relevance to the current work as it can enhance dialogue and more closely mimic human interaction to elicit engagement. Related to the dialogue examples above for richness of interaction, the illusion of intelligence can occur with a machine that can convey memory about a previous interaction by referencing content from previous dialogue or not repeating discussions that have already occurred. In addition, the machine can respond to random action, even if it is a clarifying statement, "I'm sorry did you mean…?" The illusion of intelligence can be conveyed visually (see Section 5.4.4), as well, by eliciting the experience that a machine is a human and through formalizing social discourse by employing social practices such as encouraging dialogue about a user's known interests and being sensitive about building rapport with the user before querying for personal information. The illusion of intelligence can meet many aspects of engagement, including interest, positive affect, novelty, and interactivity; however, it is reasonable to conclude that it very difficult to achieve.

Proposition six is **invisibility of the medium** [67]. Invisibility of the medium suggests that application interfaces can be designed such that the tools, displays are not at the center of the user's attention while interacting with a system. For example, a user's concern while interacting with Google assistant is that they deliver commands that fit within a structure that is understandable to the machine. Rather, the machine should accommodate the user by employing natural language processing to make the communication of the commands seamless for the user. In a driving simulator, the user should be engrossed in the simulated road events of the game, and not with the steering inputs or the mapping of those inputs. For recommender systems, users should feel as if they are interacting with the system directly to make decisions, and not concerned with the interface button operations needed to generate the options. When combined with natural voice command as the primary interaction mode, recommender systems can achieve invisibility of the medium by literally making the graphical user interface invisible to the user. Possibly the most common reason for failing to make the medium invisibility is when the usability of an application is poor. Poor usability can distract from the intended interaction experience that the application was intended to support, just as it can diminish the experience of engagement. While engaged as a result of the medium being invisible the user is not distracted by the interface itself; there is a sense of direct control over the activity.

In the next section I introduce presentation and behavioral interaction manipulations that focus on mimicking human form and behavior. These manipulations can be combined with narrative practices above to increase the likelihood of eliciting engagement from users.

5.4 Mimicking Human Interaction for Engagement

Possibly one of the most effective, but challenging approaches to eliciting engagement is facilitating interactivity through **social presence**. Bioccia et al. [68] define social presence to be the "sense of being with another." Social presence can be conveyed physically; for humans that simply means a person can be seen in person or through technological mediation, such as a webcam. Social presence is a psychological phenomenon – the "perceived presence" of someone or something else has a psychological impact on behavior. For machines, social presence can be achieved through technical means, such as textual information, images, robots, and graphical avatars. All of these give the impression of social presence, but without the need for the entity to be physically collocated – except in the case of a robot. Goffman [69] states that when communicating through a technological medium, such as text, social presence is being conveyed because it generates a sense of mutual awareness. In a conversation a human may ask a question, and a machine can convey mutual awareness by addressing the question, or it can convey awareness about the user's facial features and surroundings in the response, "In today's forecast there will cloudy skies with a chance of rain; you want to wear a bigger hat or bring an umbrella before going out." People can gain a sense of social presence because of the belief that they have access to the machine's intelligence, that is, they are able to model and make projections about the intentions of the machine. This access to machine intelligence suggests that there is an underlying intelligence behind a machine's behavior. This perception can come from interface design practices that elicit the illusion of intelligence, as discussed in Section 5.3. In the example above, a user may believe that, by suggesting a bigger hat or umbrella, a machine understands the user's desire to stay dry on potentially rainy days.

The experience of social presence can be stronger when a machine interface displays human-like features; that is, the interface is anthropomorphic. When users are experiencing anthropomorphism, they are interacting with the interface as if it is another human. Past work has shown a relationship between anthropomorphism and engagement. It is due to this relationship that makes the construct of anthropomorphism important to this work. Next, I take a deep dive on the construct of anthropomorphism and discuss how anthropomorphism can be solicited through interaction design manipulations.

5.4.1 Anthropomorphism

Anthropomorphism is a concept that describes an innate tendency for humans to interact socially with non-human objects [11], [70], [71]. Research has shown that this tendency persists despite knowing that the interaction is with a non-human object, e.g., [11]. Therefore, there is a natural predisposition toward anthropomorphism already present. With respect to developing trust, anthropomorphism can be leveraged to elicit initial trust and engagement while the user gains more experience to form better mental models of the system [72]; a more accurate mental model can improve perceived predictability of the system and calibrate trust. Past work mainly in healthcare applications has shown that designers can exploit anthropomorphism to acquire personal information from users and also to entice them into continued interaction [15], [73]–[75]. For recent empirical work has also shown that anthropomorphism has a positive impact on engagement, which has subsequent influence on users' reported intent to return to interact with a system [76]. This addresses two parts to the data solicitation problem. First, these systems must be able to coax users into providing information. Second, a recommender system must also entice the user to continue interacting with it. This can be achieved by manipulating anthropomorphic presentations.

Anthropomorphic presentations, as applied to recommender systems, focus on manipulatable elements of the user interface. That definition is the extent to which these systems have the appearance or behave like a human being [77]–[79]. Thus, the effect of anthropomorphism can be created with visual and/or verbal presentations, as well with demonstrating human cognitive and perceptual abilities that lead to the idea that the machine has intelligence [72]. Verbal presentations can be representations of human conversational structure, the use of synthetic speech and the reference to the pronoun "I" in dialogues within. Anthropomorphic cues can include images of people or synthesized speech [80]. Visual presentations can be static or animated [81], [82] depending on the desired realism. These include designing technology to mimic human form, such as rendering features of the human face, body, or skin. Visual presentations can also involve mimicking human movement. For example, researchers have animated computer models of the human face to convey human affect or create gestures by animating models of the human hand. The goal of employing anthropomorphic presentations is to create an experience analogous to interacting with another human being – an experience I will define here as ***anthropomorphism***. To elicit qualitatively high levels of anthropomorphism, researchers and designers have combined both verbal and visual presentations to create what they call a virtual representative [82], [83], a social agent [3], a relational agent [84], or an ***Embodied Conversational Agent (ECA)*** [74], [85]. In this thesis I selected the term ECA to refer to all such agents. A machine can display cognitive abilities by mimicking context understanding. This can be done by designing the machine to initiate conversation with the human based on topics that refer to the observable contextual information surrounding the interaction. For example, a machine can initiate a conversation by talking about the weather or bring up significant news events. Other cognitive abilities include speech comprehension, which enables a machine to act on voice commands that are spoken in common language with a syntax that can vary, “What’s the weather?”, or, “Rain or shine today?”. I will discuss these anthropomorphic presentation approaches in later sections.

Research has shown advantages to combining verbal and visual presentations. However, this is not true for all cases. Other studies have shown, rather, that pairing one presentation type to an application leads to more favorable outcomes than when the presentation types were combined. For example, when the objective was to increase voluntary interaction with technology, an interface that was composed of both visual and verbal presentations elicited the best results [74]. However, verbal only presentations were more effective when the objective was to solicit personal information [86], [87]. Researchers explained that anonymity is typically preferred when divulging personal information. That sense of anonymity was lost when participants were presented with a display that resembled the physical presence of another human being. It appears, based on these mixed findings, that convincing anthropomorphic presentations inherit the same advantages and disadvantages of actual human interaction. Thus, we may expect that manipulating the presentation by combining and/or subtracting different features can vary system objective outcomes accordingly.

By facilitating social engagement ECAs can elicit the perception of credibility to the system [88]. This can effectively induce trust [89], which is relevant to recommender systems because trust influences the likelihood of users accepting recommendations generated by these systems [90]. Barring other factors, such as system reliability, there is a positive relationship between trust and continued interaction with the system [40]. Other investigators have been cautious about employing anthropomorphic cueing, arguing that ECAs can be distracting [91]. Furthermore, relating to trust calibration, engagement may be hindered when users apply social stereotypes that do not align with their perceptions of the anthropomorphic cues [83]. More generally, ECAs may cause users to ascribe human capacities to the system that it does not actually possess [92], such as the expectation users have about the ability for voice recognition assistants, e.g., Alexa, Google Dot, and Cordana, to respond correctly to different ways commands are phrased and nested, “Give me my agenda”, or instead, “So what’s on the menu today, and what about my commute?” In the next section, I turn to the topic of how verbal presentations such as these can be structured to solicit information from users.

5.4.2 Verbal Presentation: Mimicking Human Conversation with Social Discourse

In 1950 Turing [93] proposed a test called the imitation game (i.e., also called the Turing test) to answer the question, “Can machines think?” The game can be set up by having a human interrogator on one side of a blind and two agents on another. One of the two agents would be a woman and the other a machine. While not knowing initially which is the woman or machine, the objective of the interrogator is to determine which is the woman. The interrogator can ask questions to build a conversation with the agents. The woman’s objective is to help the interrogator correctly identify the sex of the two agents; the machine, on the other hand will try to deceive the interrogator into making the wrong identification. The entire interaction would be in text so that the voices, of course, would not give away the game. As a test of a machine’s ability to answer questions in absence of any visual elements, the game mitigates any anthropomorphic bias by stacking the odds against the machine [94]. As a result, evidence of anthropomorphism might be considered more compelling, and what would be implied by the imitation game is that verbal communication alone can be used to elicit anthropomorphism.

Two social rules that are commonly practiced in human-to-human communication are the ***principles of reciprocity*** and ***sequence***. According to the principle of reciprocity people are generally reticent about divulging information about themselves, but an exception is made when they are the first to receive a disclosure [95]–[97]. The reticence to disclose is greater when the requests are for intimate information, because the perceived vulnerability is greater. The perceived vulnerability can be fears due to consequences that are emotional (e.g., fear of rejection when revealing interest in a high school crush), physical (e.g., fear of being mugged when revealing one’s residential address), or material (e.g., fear of fraud when revealing one’s credit card number). However, when the sequence of the requests for disclosure, from another person or computer, gradually escalates from requests for casual information (e.g., “What is your favorite color?”) to more intimate information (e.g., “Do you consume recreational drugs?”), people will be likely to follow with equally intimate responses [98]–[100]. I suggest that providing information to RSs is a form of self-disclosure. By extending the principle of reciprocity and appropriate sequence, social rules normally applied in conversations between people, can positively affect the likelihood a user would provide information to a machine.

Verbal presentations can include mimicking human-to-human conversational structure or synthesized voice [80]. This includes the use of pronouns (e.g., "I"), fictitious backstories, and initiating a conversation by divulging fictitious personal information. The general conversation strategy begins with the computer initiating with a disclosure, followed by a prompt for reciprocation with a request for information from the user (Figure 5). The conversation ends with the user either disclosing the requested information or refusing to disclose. The general conversational model can be chained to build a sequence that begins with a set of low intimacy questions before a set of high intimacy questions are delivered to gradually coax people into giving information that carry greater levels of intimacy.

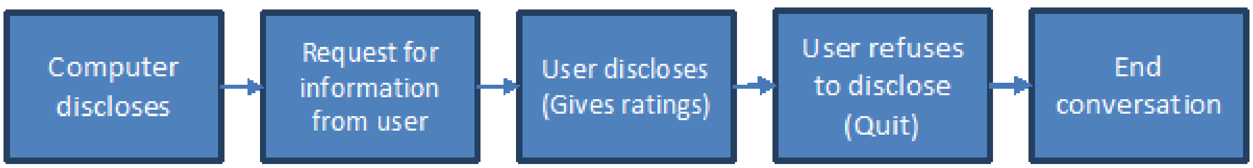


Figure 5. Basic conversational structure.

Moon [39] manipulated reciprocity to positively affect self-disclosure by conducting computer mediated interviews where each question was preceded by information about the computer. For example, "This computer has been configured to run at speeds up to 266 MHz. But 90% of computer users don't use applications that require these speeds. So, this computer rarely gets used to its full potential. What has been your biggest disappointment in life?" When the computer did not disclose information about itself the questions were direct, "What has been your biggest disappointment in life?" Similar results were found when the intimacy of the questions escalated gradually, as opposed to being delivered abruptly at the onset of the conversation.

It is important to mention that Moon was attempting to control for social cues, thereby minimizing engagement with the system in her study. In none of the conditions did the computer refer to itself as "I", nor was any of the text accompanied by graphical or pictorial representations, and the questions were phrased to be factual. That is, words that imply emotion, feelings, attitudes, or motivations were not used. Moon wanted to control for these social cues, so that only reciprocity and sequence elements of conversational strategy were examined, and to avoid inducing anthropomorphism, as she believed it would be misleading to the participants.

Bickmore, Schulman, and Yin [74], in contrast, were interested in fully inducing anthropomorphism. They asked participants to interact with an ECA that presented itself as human with a face, and a programmed dialogue that included a back story that was told in the first person. Like in Moon [39], the back stories did reveal information about the computer, but the information was fictitious because they suggested that the ECA had an experience that never existed. For example, “And I think I really developed an appreciation for exercise and being outdoors and just staying healthy and moving around all the time.” Bickmore et al., questioned the assumption that people would find a more realistic ECA misleading, and in their study asked participants if they perceived any dishonesty from their interactions with the ECA. Their findings showed that not only did the participants not perceive dishonesty from the ECA; they reported that the system was engaging and that they were likely to use the system in the future. Although Bickmore et al. did not analyze self-disclosure in their study; they suggested that the absence of mistrust in the ECA is at least an opportunity to build the prerequisite trust necessary to successfully do so. That suggestion may have germinated from an earlier study, where Bickmore and Cassell [85] explored the opportunity to build that trust by engaging users in dialogue that was not task oriented - called small talk.

Unlike with backstory, small talk did not involve telling a fictitious story about ECA’s past experiences. Bickmore and Cassell [85] defined small talk to be “any talk in which interpersonal goals are emphasized and task goals are either non-existent or de-emphasized.” An example was, “Sorry about my voice, this is some engineer’s idea of a natural sounding.” Bickmore and Cassell argued that small talk can prime users for task related conversations, as well as build the trust needed to solicit more intimate information like offering a bid for a home. Their experiment revealed that small talk had a positive influence on people’s trust only if they had a predisposition to trust, such as being an extrovert. If one was an introvert, then small talk had no influence on her/his trust in the ECA. Although small talk contained disclosures, reciprocity was not a manipulation in Bickmore and Cassell’s study. In addition, self-disclosure was not measured. Unlike in Moon [39], Bickmore and Cassell [85] provided the ECA with a human form and as a result did not isolate conversational strategy from other social cues. Thus, it remains unclear from Moon [39] and Bickmore and Cassell [85] whether there is any connection between reciprocity, the building of trust, and self-disclosure. For a discussion of the relationship between those variables we turn to a study by Zimmer et al. [101].

Zimmer et al. [101] implemented procedures similar to Moon [39], but extended the study by including an analysis of the relationship between trust in the system and the intent to disclose. Intent to disclose is whether an individual plan to provide information about her/himself and was measured using a 7-point Likert format question after visiting a website. Findings from the study revealed a positive relationship between trust and the intent to disclose. And, when users expressed intent to disclose, they followed-up with self-disclosing; i.e., measured by how frequently they divulged information upon request. Finally, consistent with Bickmore and Cassell [85], Zimmer et al. [101] found that, after receiving initial disclosure from the computer, actual disclosure was more likely to occur with participants who indicated initially that they had the intent to disclose - providing further support to the principle of reciprocity.

The degree of engagement that results from a conversation often depends on social discourse practices that can be incorporated into software agents. One such social practice is to have conversations that actively **encourage exchange of information** [102]. This can involve an agent asking questions about a person's interests and hobbies. This can include querying people's opinion and knowledge about topics that draw on commonly available information, such as news sources and popular magazines. In general, these approaches start with discussions about topics that are easy for other people to talk to. Another related social practice is to **be sensitive to one's relationship with other people** when asking questions. When two people are not familiar with each other, then the questions typically target information that is generally public. If after continued engagement the two people have established familiarity with one another from repeated interactions, then queries for personal information might be appropriate. Following these practices, a navigational recommender system might initially ask a question based on public information like "Do you know a good winery around here so I can add the information to my database?" Then, after a day of wine tasting, and repeated interactions, the recommender could ask a question based on personal information, such as "Are you intoxicated? I have identified some hotels nearby if you need to sober up." **Applying appropriate types of speech** can also encourage social discourse. As a cold-start solution, one type of speech is offering non-personal information when there is a very short or absent history of conversation between individuals; publicly available information can be shared as a conversation prompt, "Hot day today", but requests for private information as a conversation prompt should be avoided. When rapport has been established, where there is a longer conversation history and a relationship has been formed that is marked by trust (trust will be covered in Section 5.5), questions can request both non-personal and personal information. If people have a very high degree of familiarity or rapport, then ideas can be criticized. Applying this principle to a self-driving car, the vehicle recommender may recommend against going home due to traffic congestion along a route and suggest instead to go to a nearby gym to consume the delay constructively. **Expressing shared interests** is another important consideration in the design of conversational agents [102]. During a conversation, people often like to discuss and share professional interests, political views, religion and values. To encourage social discourse, people can express that they have commonalities along these dimensions, which then reinforces the first principle; have conversations that encourage information exchange. An example of this comes from social media platforms such as Facebook where user content is mined to determine user values and interests and then target political ads and products to further engage the user. Expressing common interests and values improves the chances of people liking one another and should improve the chances of people liking the agent in a recommender system. In addition, positive affect can also be enhanced through discourse that includes elements, such as flattery and jokes, and when there are displays of an agent that is physically attractive or includes child-like features.

5.4.3 Cognitive Presentation: The Illusion of Human Intelligence

Although it can be very difficult to implement, machines can be designed to perform human cognitive abilities. However, if these abilities can be achieved, even without verbal and visual presentations, the anthropomorphism can be convincing. Studying and developing how machines can accomplish human intelligence has been the primary focus of the field of artificial intelligence since its conception. Among the most prominent classes of human intelligence are spatial, verbal, visual, and social. A complete discussion of all these topics would be out of scope of the current effort and can warrant a separate dissertation. Instead, I focus on a single ability that cuts across many different aspects of human intelligence. This is the ability to process context. In human interaction, sensitivity to context allows individuals to respond appropriately to others. For machines, exhibiting context sensitivity can enhance the human experience of anthropomorphism and most important to this work, potentially lead to engagement for obtaining data from users. Also, for the purpose of this thesis, the primary goal is not for the machine to exceed the cognitive abilities of the human, but to exhibit enough to encourage more engagement. The argument is that anthropomorphism would be diminished because of apparent unnaturalness of the ability. The machine can challenge the human like another human would and still elicit anthropomorphism, but not beat the human to the extent that interaction is discouraged. To maintain engagement the human needs to be able to contribute to the interaction and gain social reward from it. One famous and recent example of this comes from the Go playing machine called AlphaGo [103] by DeepMind Technologies, which was the first computer program to defeat a human professional Go player. This has been considered a major feat for artificial intelligence because the game of Go has been more difficult for computers than chess because of the very large number of ways the pieces can be positioned on a board. As opposed to brute force computation for assessing every position possible, AlphaGo applied machine learning to limit the number of positions that needed to be processed for determining the next move by learning the human player's patterns and biases. AlphaGo is an example of context understanding [72], where the context is the latent intent of the human Go player. Initially, AlphaGo was very engaging, attracting professional players who believed strongly that they could defeat the machine. However, the story of AlphaGo would end with the human simply yielding to the machine because it was playing in ways that were not played by human players and the human could not devise a way that could improve the chances of winning. Granted AlphaGo was designed with the intent to be the first to defeat a human in the game, a more engaging version would calibrate its skill against individual human ability to encourage further engagement and make the game enjoyable and challenging. By doing so, the machine in this case, would have appropriately incorporated contextual information to encourage engagement.

In actual dialogue, context understanding can take the form of initiating a conversation by commenting about the weather or even mining an individual's social media data to present topics of interest. For example, a smart phone can verbally recommend "rainy day jazz" knowing that the user has been at a coffee shop for more than 15 minutes and given the appropriate weather outside. As a more basic example, iPhones today will know if a guest is trying to access a private network and then prompt someone who is adjacent and has access to send access credentials to the guest. The ability to process and recognize artifacts in the user's surroundings can give the impression of visual intelligence. A machine can display this by commenting on graphical elements of someone's T-shirt or, if the appendages and eye features are rendered, point or gaze at an object that a user is referring to in dialogue. Achieving human intelligence, especially social intelligence, as part of a machine's capabilities has been a major challenge, with convincing implementations that are being realized only recently with machine learning. Consequently, any near-term research relying on mimicking machine intelligence would actually require humans to stand-in for the machine while participants interact with graphical renderings of a human to ironically convince them that it is a machine [104]. Interaction between humans provides a sense of serendipity to the interaction, as well as context. When machines fail to achieve this by exhibiting repetitive dialog and motion, the perceived cognitive intelligence of the machine diminishes along with the anthropomorphic experience [105]. While cognitive intelligence, if successful, can elicit a stronger anthropomorphic experience, visual presentation does not need to be so realistic to achieve engagement. In some cases, a minimalist human-like rendition can elicit more engagement than near realistic implementations. However, if and when the machine is able to exceed human cognitive abilities, the anthropomorphic experience can be diminished and discourage engagement as there is no reward outside of practical utility for engaging the machine. Yet this is not an argument for limiting the machine abilities to the level of the average human. There are humans that are not average, but like these humans, the machine may need to balance their exceptional abilities with convincing verbal and visual presentation to preserve engagement with other people. I discuss visual presentation further in the next section.

5.4.4 Visual Presentation: Behavioral and Form Realism

In the present section I review literature relevant to the application of visual anthropomorphic cueing in eliciting self-disclosure. The overall consensus is that people's willingness to provide information diminishes in the presence of visual anthropomorphic cues alone [86], [87], [106]. However, when visual cues are combined with verbal presentations, i.e., conversational strategy, the additive effect increases the likelihood of users providing information to the system.

The literature has identified two classes of visual anthropomorphic cues: 1) form realism [81], [82]; 2) behavioral realism [87], [106]. ***Form realism*** is the extent to which facial features and skin textures match those of an actual human being. ***Behavioral realism*** is how closely a computer rendering is to the actual movements that create expressions on a representation of a face or body. Visual presentations can vary across these two dimensions, and in different combinations of high and low form and behavioral realism. The illustration in Figure 4 shows some examples that combine form and behavioral realism to create increasingly convincing human analogues.

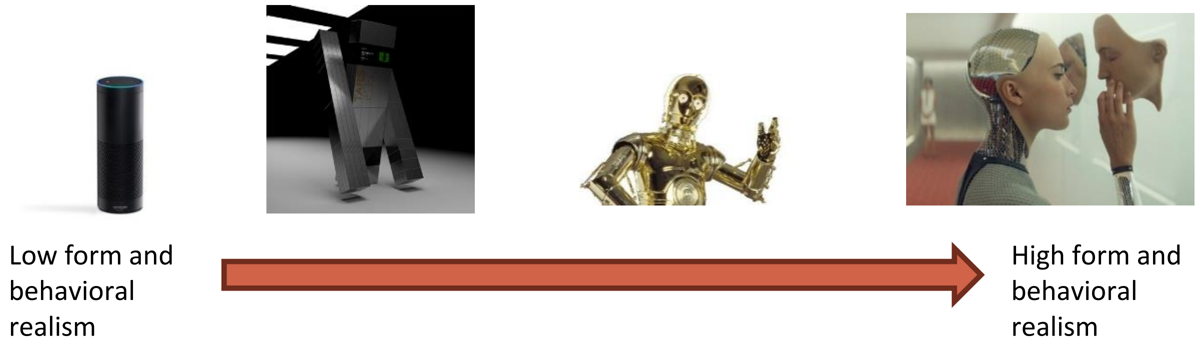


Figure 4. Examples of increasing form and behavioral realism.

In general, people divulge more intimate information in the presence of fewer visual anthropomorphic cues. Researchers have explained the reason is because people prefer anonymity when disclosing, particularly when the information they are divulging is intimate. Thus, when an ECA possesses features such as eyes and ears like that of a human being that anonymity is taken away. Several findings support this. In Kang and Gratch [106], participants were asked to disclose to avatars that varied in the realism of their depiction. At the highest level of realism, the avatar was a raw video stream of a hidden interviewer. The mid-level rendition was a degraded black and white version of the original video, in which the textures and graphical details remained. At the lowest level of realism, the video had an abstract rendering with much of the original skin texture removed. Using the Altman and Taylor self-disclosure classification system, Kang and Gratch discriminated between high, medium, and low levels of participant disclosure, and found the lowest scores with greater ECA realism. Thus, when comparing text-only interfaces against talking-faces, the idea that people will self-disclose more frequently with fewer anthropomorphic cues should continue to hold, as was indeed shown by Sproull et al. [86].

Using the text-only and talking-face interfaces Sproull et al. [86] delivered questions from a psychological scale, such as the Texas Social Behavior Inventory of Self-Worth, which asks questions like, "When I am in a group of people, I have trouble thinking of the right thing to say." Participants responded true or false, or on a Likert scale that could range from "strongly disagree" to "strongly agree." In addition, they asked open ended questions such as, "Tell me something about yourself," using an ECA in text-only or talking-face conditions. It is important to note here that Sproull et al. did not indicate the use of any conversational strategy when participants interacted with the ECA. That is, the principle of reciprocity [39] and the gradual increase in intimacy as discussed in the previous section, were not applied. Instead, the participants controlled the pace of the interaction and the questions were direct. For example, "How relaxed do you feel?" Participants had the option of replying to a question or skipping it. To proceed they clicked "go ahead" on the screen. From their analysis, Sproull et al. found fewer missed questions overall with the text-only condition, suggesting that participants were more willing to disclose when visual anthropomorphic cues were absent. Given Sproull et al. along with the other studies reviewed up to this point, it appears that the presence of visual anthropomorphic cues generally suppresses self-disclosure. However, evidence from Bailenson et al. [87] paint a slightly more complex picture. In their study, the mode in which the questions were presented not only affected how much information was divulged, but it changed how it was divulged.

Bailenson et al. [87] measured verbal and non-verbal self-disclosure in response to 30 questions that were rated equal in intimacy. Verbal disclosures were identified using two coders blind to the experiment conditions. Nonverbal self-disclosures were identified using a face tracking algorithm that counted movement from 22 points on the face known to vary with expression and reflected emotional responses to the questions. Using these measures Bailenson et al. compared self-disclosure across three ECA conditions: 1) voice-only; 2) synthetic face; 3) raw video. The synthetic face condition provided behavioral realism, where the movements of the face were realistic. However, form realism was absent. Form realism was the extent to which the facial features and skin textures were real, such as in a photograph. The raw video condition provided both behavior and form realism. Consistent with past research, Bailenson et al. found that verbal self-disclosure diminished with ECA realism. Verbal self-disclosure was least with raw video, and better for both synthetic face and voice only, which in turn did not differ. Nonverbal disclosures, in contrast, diminished in the presence of a face, regardless of the realism. This supports the notion that in the presence of facial features, people become reluctant to self-disclose due to a reduced perceived anonymity.

The studies reviewed in the previous section demonstrated that if the appropriate conversation strategy was applied, people can be coaxed into intimate self-disclosure. Those findings were made in situations that specifically controlled for visual anthropomorphic cues. The researchers did this because they were concerned about potentially adverse effects, such as creating mistrust in a system that was attempting to pass as a human. Nevertheless, realistic ECAs can be utilized to have desirable impacts, such as increasing the frequency of health related behaviors and improving learning in teaching contexts [15], [74], [85], [92].

Kang and Gratch [106] believed that by employing visual anthropomorphic cues along with the appropriate conversational strategy, one can take advantage of the benefits from maximized realism as well as achieve intimate self-disclosure. In their experiment they structured conversations so that the reciprocity principle was applied by having the ECA initiate with a self-disclosure before requesting information about the participant. The ECA disclosure varied between high, low, and none, and the condition was crossed with behavioral realism. The ECA behavioral realism was restricted by limiting head and facial animation, which varied between high, low, and audio-only. Contrary to previous findings, Kang and Gratch found greater self-disclosure in the presence of a realistic ECA, but it was also necessary that the ECA initiated with high intimate self-disclosure. Thus, by employing the appropriate conversational strategy, intimate self-disclosure can be achieved with a high degree of ECA realism.

ECAs provide a vehicle for eliciting anthropomorphism. ECAs can be presented visually in humanoid form or verbally via text-only interaction or synthesized voice. Past research has found that visual representations can improve the quality of user experience. However, when soliciting information from users was the primary goal of an interface, there was more support for employing verbal ECA representations. The success with verbal representations can be attributed to conversational strategies that incorporate reciprocity and gradual requests for increasingly intimate information. When combined with a visual form and an appropriate conversational strategy, it is possible that an anthropomorphic recommender system can successfully solicit information from users. Empirical work will be needed to investigate this notion.

In addition to the social discourse manipulations above, trust can have an influence on whether users will continue interacting with a system in joint activity and whether they will experience engagement. This is a presumption I am making based on past work investigating trust in recommender systems. In this thesis, trust is being treated as a result of a user’s experience with an application. For recommender systems, trust can be shaped over the course of interacting with a system but may ultimately be informed by the relevancy and acceptability of the options they generate. In the next section, I close this chapter with a review of the literature on trust and follow with a discussion about transparency – a principal determinant of trust.

5.5 Trust

As discussed in Chapter 2, user data plays a significant role in improving recommender system output (Figure 6). One indication of the quality of recommender system output is user acceptance. User acceptance of recommender system results can also lead to improved user trust in the system, e.g., [17], [18]. Trust is important to this work because of its potential to impact two principle constructs for obtaining user data; these are joint activity and engagement. The expectation is that trust will provide two paths into an interaction loop that can result in more user data. The main path goes directly into joint activity where information exchanges with the recommender system can result in more user data. The alternate path is by augmenting an engaging experience so that the user would be enticed into joint activity with the system. The research investigating the direct relationship between engagement and trust is sparse, and it is unclear if the findings are relevant to this work because of the disparity in the way engagement is defined across the different application areas in which these investigations are being made. Intrinsically, joint activity involves multiple actors working together, and so intuitively one would expect trust to be a factor in such interactions. However, the opportunity to investigate trust in joint activity with machines has only recently been given serious consideration because of advancements in automation has enabled machines to exercise many of the same capabilities that humans have been able to exercise in joint activity with other humans. I reserve further discussion about human machine joint activity to Chapter 6. In this section, I focus on defining trust and close with a discussion about one principal determinant of trust – transparency.

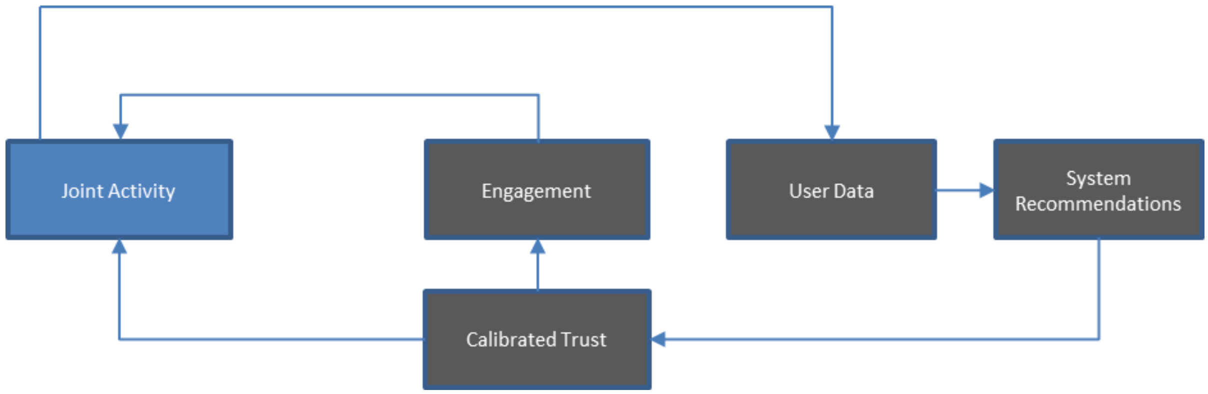


Figure 6. A framework for trust, joint activity and engagement.

Although there are many definitions of trust, I will apply the following from the seminal paper by Lee and See [107], "**Trust** is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability." Trust is voluntary and development of trust takes time. Over the course of interacting with a system, users learn about the capabilities of the technology and calibrate their trust accordingly [42]. When a system is able to meet a user's expected performance criteria and goals, improved trust is typically the outcome [107]. Recommender systems provide a capability that users are not able to achieve. That is to process large quantities of data to generate a more manageable set of options for decision-making. When users employ a recommender system to accomplish a task that they are not able to supervise, they have elected to make themselves vulnerable to the system and are conveying trust in the recommender system to assist with accomplishing their goals. A user can continue to rely on a system to accomplish a task, but this is not necessarily indicative of engagement. Reliance is interaction with a system where a user elects to interact with a system for a task because other options are not feasible. For recommender systems, the options are to shift through volumes of data manually or have the automation do it. Uncertainty and vulnerability are present with reliance, but a user does not need to trust a system to rely on it. Reliance on a system or application can be detrimental because it sets the user up for inappropriate use of the technology. If a recommender system does not have sufficient data to provide users with the best set of options possible, users will always be left with choosing among bad options – the goal of supporting decision-making would not be achieved. As a user gains more exposure to a system and learns about its capabilities and limits, they may be able to better adjust their level of trust in the system. When information about the capabilities of the system is insufficient, the user is then susceptible to trust biases. I discuss these biases next.

Certain important biases can emerge if trust is based on early exposure to a system and there has not been enough time to observe its overall reliability. That is, when it is insufficiently calibrated. One such bias is overtrust. **Overtrust** occurs when a person trusts the system to do more than it can do. In many cases this type of bias can have serious ramifications. For example, the autopilot driver assistance automation in Tesla vehicles allows limited vehicle self-driving for highway conditions. As part of the normal operations of the vehicle, drivers are expected to keep their hands on the wheel and monitor for road events that would require them to return to manual driving. Overtrust in the autopilot can be observed when drivers fail to follow this requirement by disengaging from the vehicle operation and choosing to engage in secondary tasks such as operation of smart phones. These drivers have become desensitized to the fact that the autopilot functionality is limited.

In contrast, distrust occurs when trust falls short of the actual capabilities of the system. For recommender systems this is often the case if the system fails to provide acceptable options during initial interactions. The distrust bias is observed when the user disengages the RS altogether, thus denying the system additional opportunities to obtain user data and improve recommendations. Recommender systems apply filters to process very large sets of options, as users lack the capacity to perform the task independently. RSs must capture initial engagement early in the process to spool-up on information from the user so that it can continue to improve the options it delivers over the course of the interaction. However, if trust is lost in early interactions with a system then diminished engagement can follow, resulting in the failure to obtain enough data to improve recommendations. Loss of trust can occur for several reasons. With recommender systems, lack of data from users can lead to poor recommendations. When the recommendations are poor, the system is not achieving the goals of the user in making decisions making it untrustworthy to the user. Loss of trust can also occur if the design goals of the recommender system do not align with the user goals. If for example, designers of a recommender system rig the solution set to direct a user’s options, as opposed to produce a set that best represents the user’s goals, trust would diminish if the user were to discover the ploy. When an RS interface is conversational, users can learn to calibrate their trust so that it reflects actual system capacities by “tweaking” the inputs to see the impact on the recommendations presented [24]. This allows the users to get immediate feedback on their interactions and more control with the system – attributes of engagement. For recommender systems, trust can be informed by both its ability to generate recommendations, as well as the quality of those recommendations. For the former, an example would be a driving route that either fails or succeeds to generate directions for a trip. For the latter, a recommender system for recreational travel may prioritize options for points of interests within a geographical location. Being able to interact with a system in trials to evaluate the outputs against user inputs calibrates trust. **Trust calibration** [108] occurs when the user trust in the system reflects the capabilities and limitations of the system. When a user exhibits calibrated trust, they will likely employ a system for tasks that it is actually capable of accomplishing, and not employing it for tasks that are outside of its limits. As the relationship between inputs and the recommendations along with the information about the systems actual capabilities are revealed to the user, they explain the underlying reasons for those recommendations – this is a form of transparency.

Transparency is information that explains a systems behavior [24]. According to Seong and Bisantz [109] an automated system is transparent when the inner workings or logic are known to the human and they are able to understand the system as a result. Chen et al. [110] further specify that transparency comes from the ability of an agent to convey intent, performance, future plans, and reasoning process to a user. For recommender systems, the latent information that is revealed through transparency describes the selection or rejection of an item that was not explicitly requested, "we are recommending these shoes to you because other people who shared similar preferences bought them." Explanations like these are presented in common language and not in the logic behind matching user profiles, because another aspect of transparency is that it must be understandable to the user. For recommender systems, the overall purpose would be defeated if all the underlying reasons for a system's functioning were completely exposed. According to Chen et al. [110], automation transparency is "…the descriptive quality of an interface pertaining to its abilities to afford an operator comprehension about an intelligent agent's intent, performance, future plans, and reasoning process." The key aspect of this definition is that explanations are comprehensible to the operator.

Introducing transparency can help to address overtrust and distrust by making information about the system behavior available to the user. This information can be provided in advance so that the user can then make predictions about the behavior of the system, such as the filtering categories in conversational recommenders [24]. In the Tesla vehicle example, this could mean marketing the capability as a "driver assistant", as opposed to "autopilot." Explanations for the recommender system results after they have been generated can also provide transparency if they are presented in a way that is understandable to the user. Car navigation applications do this by informing users when they are on the fastest route possible, or when there is a traffic event such as a car accident that has prompted the automation to propose a reroute. Making information that explains the results of the recommender system in this way improves the likelihood that users will align their trust against the actual capabilities of the system – this is the trust calibration discussed previously [107]. With calibrated trust, the likelihood of users engaging the recommender system can continue to improve, thus creating more opportunities to obtain data and further refine options for users.

In conversational RSs, trust and transparency is built through ongoing exchange of information with the user. These two factors can potentially facilitate engagement, creating opportunities for the system to obtain more information. The process of paring down options by conversing with the RS is a joint activity and the context information emerges from this activity. Early in the process there is increased risk to a user abandoning the system if it fails to achieve trust by producing poor recommendations. When an RS has very little information about a user, the likelihood of generating poor recommendations is high. This is at the heart of the data solicitation problem. To resolve this conundrum, a solution should consider options for engaging a user for reasons outside of functional joint activity. Recall that joint activity can also occur for recreation. In the next section, I discuss how an already present tendency for users to socialize technology can be manipulated to elicit social engagement. This tendency can be enhanced through visual and verbal interface implementation approaches that included eliciting anthropomorphism and the structure of the conversation with the machine. The notion is that if a user engages an RS for social reward, i.e., enjoyment, they would be less sensitive to poor initial performance of the system and allow the system more opportunities to spool up data from the user and subsequently improve recommendation performance.

In the next chapter I continue this discussion by addressing engagement for the purpose of accomplishing physical and cognitive work, or practical activity. Like form and behavioral realism, a machine can exhibit functional realism by displaying human knowledge and skills, i.e.

capacity. When machines exhibit human capacity, the potential for greater engagement through joint activity can result due to the improved utility of the machine. Researchers have developed new frameworks that describe how machines engage in joint action with humans. The extension here is that these frameworks can be used to guide the development of RS interfaces to extend joint activity with users and obtain data from them during information exchanges. Before I discuss these frameworks, I present some background related to prerequisite work in the basic allocation of functions between humans and automation.

6. THE EMERGENCE OF HUMAN MACHINE TEAMS

6.1 Levels of Automation (LOA)

In this chapter I review the literature in human automation interaction because it serves as a starting point for describing how functions can be shared in human machine teams, and then how teams can lead to more engaging interactions for improving the likelihood that data can be obtained from humans for better RS output. I begin with defining function and what it is to automate. The definitions I present are not absolute, but they are useful for facilitating this discussion.

Here I define a function in absolute terms for the convenience of discussion, however, the term has been used more broadly to describe both goal orient actions as well as available capabilities and roles [37], [111]–[113]. For this document I define a function as the most elementary concept of an operation [114], For example, in a construction of a building, a function can be site excavation. The excavation may then require individual tasks, such as the removal and then displacement of soil; thus, when a function is performed, it will include all the tasks needed to perform that function. A function is automated when a machine performs the function in place of a human [115]; this will be definition applied for automation herein. When the reallocation of that function becomes fixed to a machine, so that a human no longer performs the function, then it is simply a machine operation – not automated. However, if both machine and human continue to perform a function, such as dish washing, then the function remains automated. This definition captures the way an entire generation of researchers have viewed automation – that it is a set of functions that are performed independently by either human or machine. This function allocation strategy originates from the "men are better – machines are better at" (MABA-MABA) framework that was introduced by Fitts [116].

The MABA-MABA framework was developed when machines were built primarily to perform difficult physical work or brute force calculations. These were functions that humans, with exception to a gifted few, were performing at a cost to quality and accuracy. When advances in computing turned isolated functions into groups of complex automated functions embodied in the hardware, automation became an object. As a machine, automation has become far more capable. Unconfined to single behaviors, advanced automation has left humans with functions that do not accommodate for the limits of the human, such as system monitoring [117]. This is often a product of designers inserting machines to take a task without recognizing that it has an impact on the overall work dynamic and output of the human – who is employing the automation to assist with a job. Advances in automation have created an opportunity to exploit automation for coordinated action with humans. That concept, introduced by Sheridan and Verplank [118], influenced a generation of research in automation. To support the concept, they presented the Levels of Automation (LOA) framework that lays out the human automation interaction space, so that empirical investigators can identify any corresponding impact on human performance and inform system designers on how to exploit automation for human/machine coordination.

LOA is a discrete approach to describing the coordinated action between humans and automation [118]. Sheridan and Verplank described 10 LOAs in their framework. In general, the levels of automation are manual (Level 1), supervisory (Levels 2 to 7), and automated (Levels 8 to 10). At the manual level, there is no automation present, hence no human automation coordination, and the human is left to execute the task. At the supervisory level, the human oversees the behavior of automation and intervenes at their discretion. At the fully automated level, responsibility for executing a task is then shifted fully to automation, while information about the actions performed by the automation is made available to the human. Levels 2 to 3 represent the interaction strategies employed by RSs today. At Level 2, the computer presents options. At Level 3, the computer will present options and prioritize one. Although the original LOA provided 10 levels, the intent was for greater granularities of automation to be defined within these levels [119], however, the allocation of functions remains discrete; the human is doing the task, or the machine is doing the task – not both. There is no exchange of information; any latent contextual information remains internal to the human and manifest as high-level direction to the automation. Recommendations provided by the machine in the LOA framework are informed only by observable data. Level 8 to 10 automation allows the automation to be free of supervision and the capabilities enable adaptive automation strategies that leverage a machine’s ability to incorporate contextual data about the human and system state, e.g., recovering an aircraft when a critical altitude is met, and human incapacitation is detected. In this framework, humans are responsible for executing a function, but automation has the authority to usurp execution authority[4] if/when it detects a breakdown in human performance [120], as well as return execution authority to the human when performance recovers. Although the adaptive automation scheme incorporates contextual data, it follows the tradition of discrete function allocation.

From a phenomenological perspective, to elicit engagement for obtaining data from the human, a human automation interaction framework will need to accommodate for joint activity through function sharing between the human and machine. The challenge here is that the technology that will enable a machine to perform many of the high level human cognitive functions has only recently emerged. Correspondingly, researchers are engaging in early discussions about new human and machine interaction frameworks, as well. In the next chapter I report some of the recent developments in computing that will enable machines to interact with human teams, and discuss how these machine capabilities, when fully realized, have the potential to achieve the human engagement necessary to address the data solicitation problem.

In the previous chapter I presented the MABA-MABA function allocation strategy and the LOA framework. In both concepts, functions are allocated exclusively to either the human or machine. In part, this can be due to a lack of capacity on the part of the automation, leading it to be pigeonholed for specific functions. In systems with poor transparency, humans drift out of the loop, leaving them with little awareness about the behavior of automation. This has negative ramifications for data solicitation because the user is disengaged even before the first prompt for information. Recently introduced human machine teaming frameworks offer some remedy to this issue. They center around joint action, but to achieve joint action that is analogous to human-human cooperation will impose more demanding capacity on automation, which include: the ability to communicate between agents and the environment; independently perform all the functions of a human role; and exercise self-governance. These expanded capacities describe what the literature has termed autonomy.

[4] This is not the only characterization of adaptive automation. Other definitions include dynamic changes in level of automation by either human or automation [164] and dynamic allocation of functions to adapt to changes in workload and/or operational demands.

6.2 Autonomy

For this thesis, I adopt the notion that autonomy is a state [121], [122]. Both machines and humans can exhibit autonomy; however, as it is a social construct, it is convenient that humans serve as the model. Although this may seem self-serving, it is an important distinction because the human machine teaming frameworks I discuss in the next section aim to treat machines as human analogues, with human capacities. When present capacities allow it within a current context, autonomy can be exercised. In other cases, the capacity may not meet the demands of the operational environment, then independence cannot be exercised. If there remains a desire to complete a task then the entity, machine or human, will need to form an interdependent relationship where skills can be complemented to complete a task.

When a human or machine is exhibiting autonomy the following characteristics are being exercised: viability; independence; and self-governance [123]. Viability is the extent to which an agent is robust to changes in an operating environment, such as a robot that can transit across both rocky and level terrain. To achieve viability an agent may need to retool itself or self-repair so that it can take appropriate action when met with an unplanned challenge; this segues into the next characteristic – independence. For the most part, the definition follows Johnson et al. [44]. That is, an agent possesses the full capacity to perform an activity within a given context; the only distinction here is that Johnson et al. specifies that the agent does not require monitoring from a human. The monitoring may be desired for providing transparency to a system and benefits the human, but an independent machine will not need ongoing corrective input or freeze into inaction when it reaches certain limitations[5]. A machine not receiving direction from a human must exercise self-governance. A self-governing machine would then be able to take responsibility for mission goals and control of resources [123]. If these characteristics were to hold as requirements for autonomy, there would be no machine that qualifies today. However, as designers advance the state of the art in automation, it would be helpful to implementers to know what capacities an autonomous machine can achieve – even though it may not be full autonomy. It may be unattainable to achieve full autonomy where autonomy can be exercised all the time; context may place limits on when those capacities can be exercised, and this may apply for both engineered and

[5] An independent machine may break when information is lacking or if predetermined goals do not meet changing demands from the operational environment.

natural systems. For designers, it may be useful to classify autonomous systems by their available functions for an appropriate mapping to the operational context. To that end, NASA has proposed such a taxonomy.

NASA's autonomous system taxonomy[6] is composed of four basic functions [124]: situation awareness; reasoning and acting; engineering and integrity; and collaboration and interaction. A machine that is situationally aware has the capacity to interrogate, identify, and evaluate the state of the environment and itself. Reasoning and acting are achieved when an agent can analyze and evaluate situations to make decisions, and self-direct to achieve a goal or mission. Unlike in Kaber's definition, being able to define a goal and mission does not need to be within the capacity of the machine and can be provided by a human, relaxing the criteria for autonomy so that current systems can at least be considered for joint activity with humans. The design and development effort for autonomous systems is reflected in engineering and integrity, where efforts like verification and validation, testing and evaluation of the system, as well as operational assurance take place. The item most relevant to this thesis is collaboration and interaction. A machine that is capable of collaboration and interaction can share knowledge and understanding[7] with other actors, identify the intent and behavior of other actors, negotiate goals and tasks, as well as build trust. Situation awareness, reasoning and acting, and collaboration and interaction are common themes in human machine teaming frameworks; all which center on joint activity afforded by machine autonomy. In the next section I provide a selective review of representative human automation teaming frameworks.

6.3 Human Machine Teaming Frameworks: A Few Considerations

The most cited definitions for team have two commonalities. The first is that teams are composed of two or more actors and the second is that they are engaged in activities to achieve a common goal [125], [126]. The element that appears to vary between definitions is how the activities are carried out within the team. This can be independent, but coordinated action based on a common script [45]; a football playbook is an example. Alternatively, a team can be jointly active, such as

[6] This is for engineered as opposed to biological systems.

[7] The term understanding applied here refers to the ability to interpret information with respect to the context in which it is being presented.

that performed by a horse and rider, where the horse has the agency to remain on path while the rider can provide further directional cues. In this chapter I provide a selective review of models for human machine teaming where coordination from a script and joint action are treated as options that a team can employ to accommodate for the demands of a task. Current models of human machine teaming are aimed at informing practical engagement for accomplishing work, and although the focus of this work is on social engagement to obtain data, I review these models[8] to identify the gap.

Current models of human machine teaming are aimed at designing the team structure such that machines can be more proactive with interacting with a human. The presumption here is that changing the team dynamic so that the machine can behave more like a human, as opposed to only receiving commands from the human, would encourage joint activity with the machine and lead to more information exchanges where user data can be obtained.

6.3.1 Theoretical Frameworks

Recall from Clark [43] that joint activity includes goal-oriented activities, as well as recreational activities; both requiring communication for exchange of information between actors. Current machine teaming frameworks only consider the former. This seems reasonable because they meet the immediate needs of implementers for applications such as robotic undersea and planetary teleoperations. The collaborative control model (CCM) was originally developed for these applications [127]. Consistent with the definition of autonomy presented previously, CCM views the machine as a team partner as opposed to a tool. The machine in CCM can request approval from the human before acting, but it is not required to do so. In CCM, joint activity is performed by designing the machine to negotiate control when it encounters situations that requires transitions between autonomy and dependency. The machine is also able to query the human to close any gaps in its own capacity, such as validating obstacles, a characteristic of interdependency. Finally, to facilitate communication aspects of joint activity, CCM proposes that the machine and human exercise flexibility in the format of the information being exchanged. For example, machine proprioception may detect rough terrain and convey that verbally or graphically and the human

[8] The literature applies the use of the term model to describe both computational models, as well as frameworks, which offer principles and guidelines as opposed to detailed methodology. The work selected for review here was determined by whether I thought the information would be actionable for a designer.

may provide inputs using verbal or command prompts. This was demonstrated in a design implementation based on CCM by Fong et al., who employed natural language prompts to engage the user, "How dangerous is this object?" Although verbal anthropomorphism can be elicited from this example, the purpose was to facilitate task oriented joint activity; the CCM framework itself does not formalize the structure and content of the conversation for social engagement. If the tendency to socialize technology was present in the Fong et al. robot implementation, it would have been a convenient byproduct of the ongoing task-oriented activity.

The coactive design framework [44] is an extension of the collaborative control model (Figure 7), and stresses agent interdependence. Thus, the first major difference between the models is that the machine and human are both agents. Recall from Fong et al. [127], it appears as if the machine is engaging in ongoing activity and then the human is brought in only to close the gap when the machine encounters a lack of capacity to deal with a new challenge. In coactive design, the machine can be brought in to support the human. The second difference is that the cognitive processes for the human as well as the machine are represented in the model. The human processes are non-linear, e.g., sensing, beliefs, perception, planning, decision-making, and acting. For example, where a machine may only decide after the necessary information about functional goals, needs, and objectives have been determined, a human can make a trip to a supermarket with a shopping list, but then leave with several serendipitous purchases beyond the list. However, the machine processes, as suggested in the example above, are linear, e.g., sense, interpret, plan, decide, and finally act. To support joint activity and mitigate the differences in processes between humans and machines, the coactive design framework proposes that an interface needs to support observability, predictability, and directability. Observability is about making information from one agent available to other agents. That information can include status, e.g., agent is working versus recharging, knowledge about collective team capacity, task, and environment, i.e., contextual information, observable to actors in the team. Predictability means that the goals and intentions are observable so that one actor's actions are predictable to other team members and thus they can adjust their own actions appropriately. This enables synchronization of joint activity. Directability is the sharing of responsibility among agents for self-assigning roles or assigning roles to other agents; this also includes issuing commands and allocation of tasks. This leverages any autonomy that can be exercised by team constituents. Observability, predictability, and directability can be set as goals for the design of a human machine teaming interface, but there needs to be further formalization of the principles. For example, the information exchange required for observability can be modeled on conversational interaction techniques for recommender systems.

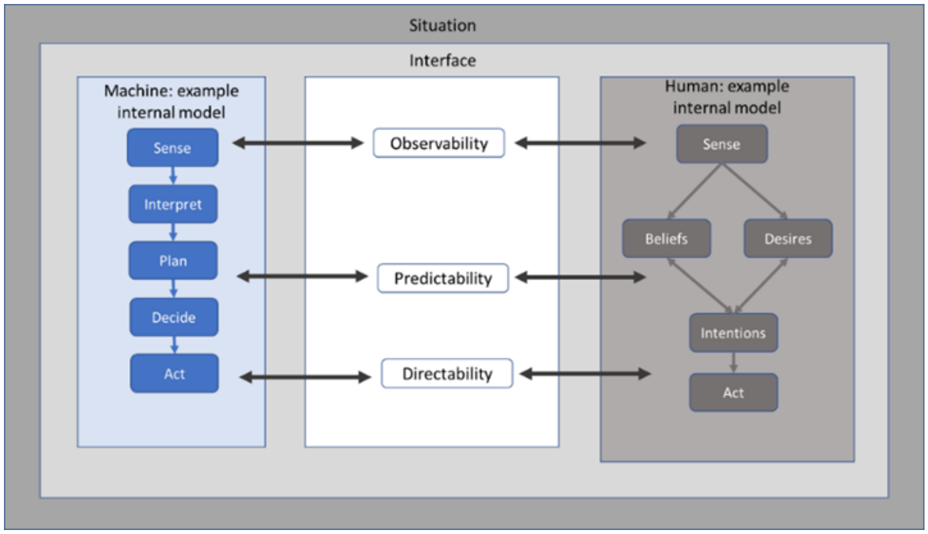


Figure 7. Collaborative control model. Adapted from Johnson et al. [44].

The mixed-initiative interaction framework centers on directability, emphasizing joint activity, where agent roles are not predetermined[9] [128]. The interdependence is instead opportunistic, leaving the agents to determine the roles based on the capacities required to solve a problem or complete a task. The agents are not required to work together for the entire duration of the joint activity. They can work independently and then return to work jointly as required by the activity as it unfolds. A joint activity will contain embedded sets of actions with individual start and end times. The coordination effort that is exercised by a team involves synchronizing both independent and interdependent actions within the joint activity. The mixed-initiative interaction framework shares a level of abstraction common with many other frameworks. It offers some insightful philosophy into the design of human machine interaction, pointing to the dynamic trading of control authority between human and machine agents as the best way to exploit autonomous systems. Next, I turn to describing practical frameworks where some of the principles described in this section may apply.

[9] In CCM and Coactive design frameworks the roles can be predetermined, but in mixed-initiative the roles are never predetermined.

6.3.2 Practical Frameworks

Aimed at formalizing the design of human-machine cooperation structures, Pacaux-Lemoine and Vanderhaegen [129] defined an interaction space that focuses on a decomposition of interdependent relationships. The model characterizes agents in terms of capacity and ability to cooperate [10], agnostic of whether the agent is human or machine [130]. As previously defined, capacity is an agent's skills and knowledge; Pacaux-Lemoine and Vanderhaegen [129] extend the definition to include not only existing skills and knowledge, but also the perceptual ability to acquire, through communication and observation, knowledge that the agent did not originally possess. The ability to cooperate is the agent's ability to communicate and form models of another agent. A convenient example of a machine's model for humans comes from the user profile generated by collaborative filtering techniques and used to predict preference. As drivers, humans form models for other actors, such as cyclists and other drivers, based on their own experience. A driver who is an avid cyclist will be able to tailor expected behavior to the observable skill of another cyclist and be able to negotiate maneuvers with the cyclist indirectly. As illustrated in the previous example, being able to form models of other agents affords the ability to make projections about their behavior, even when communication is absent.

Pacaux-Lemoine and Vanderhaegen [129] were interested in identifying how much communication needs to occur as result of the demands imposed by the task environment and the individual capacities of the agents. Similar to the definition of interdependent relationships by Johnson et al. [44] for independent agents, Pacaux-Lemoine et al. [129] emphasized forming teams of independent agents for enhancing task performance. Forming interdependent relationships in this framework does not need to be motivated by the lack of capacity from dependent agents. Independent agents can form interdependent relationships to additively contribute individual capacity to enhance the efficiency of teams by increasing the speed of performing a task or accomplishing a greater number of tasks. In a well-practiced doubles tennis team, the two players carry very well-defined and accurately mental models of each other's playing patterns and style. This allows them to make accurate projections about where they will move across the court, while reacting quickly to changing dynamics of the game without needing to openly communicate at

[10] The terminology applied here is not that used by Pacaux-Lemoine et al. (2013), but equivalent, and maintains consistent use of terms within the document to avoid confusion. Capacity was used in the original publication to describe workload and attentional limits, not skills and knowledge.

every play. Tennis players are independent actors, able to play singles matches, but they can form these interdependent relationships in doubles matches to quickly cover more area of the tennis court[11]. Although tennis players in a doubles team form predetermined mental models through practice, they can also update the models through passive observation as the team modifies behavior in response to unexpected plays from the opponents. If the team sees an opponent approach the net, then both players may return to the rear of the court as a practice strategy that is prompted by specific actions taken by the opponent. For engineering, coordination models like these can be applied to inform decisions about when agents should be independent, and when more interaction must occur between agents in a team for coordination. When mental models are not accurate, e.g., a tennis team that has not practiced together, and agents are dependent, e.g., tennis teams composed of neophytes, an interaction scheme that involves more communication might be needed for proactive coordination. In a team of agents that have access to common sources of information and well-formed mental models of behavior, such as tennis teams that practice together regularly, the agents can passively coordinate their own actions against the predicted actions of their teammate. In the tennis example, if a player knows that the partner tends to play close to the net and then observes that partner approaching the net, then that player may play towards the back of the court to cover area lost to the aggressive net play. The player attacking the net can coordinate his/her own actions, knowing that the teammate will have the back of the court covered. Pacaux-Lemoine et al., presented this as a general approach to deciding between various coordination strategies based on goals and objectives, as well as constraints inherent to each of those strategies. Next, I review a computational methodology for informing human machine teaming designs that takes such constraints into consideration and provides an ability to simulate, predict, and assess the performance of various teaming structures in advance of actual implementation.

IJtsma et al. [37] proposed a methodology for defining and evaluating team structures. Like the model proposed by Pacaux-Lemoine and Vanderhaegen [129], a team structure was characterized by how much coordination activity needs to occur between independent agents. These structures are determined by system objectives that describe what mode of interaction between the agents are

[11] Tennis doubles are also an artifact of sports organizations trying to make the game more interesting and marketable.

best or selected. The methodology first decomposes functions into three hierarchical categories with functional purpose at the top, then abstract functions, followed by generalized functions; and physical functions (Figure 8).

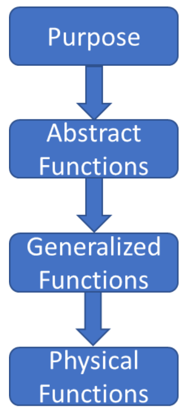


Figure 8. Function hierarchy. Adapted from IJtsma et al. [28].

An example of a functional purpose would be unmanned aircraft system (UAS) operations designed to deliver vaccines. Common abstract functions are safety and efficiency. Generalized functions can be to operate aircraft, package vaccines, and vehicle maintenance. Physical functions further decompose general functions into actions. For vehicle maintenance, the physical function can be to charge batteries and replace propellers. For preparing vaccines, this can include retrieval from storage and packaging. Each physical function will be constrained by physical resources and require information resources. For example, physical resources can be batteries, chargers, screw drivers, and insulated bags and cargo boxes for vaccines. Some examples of information resources are descriptions of weather conditions, battery capacity, number of UASs in operation, vaccine inventory, and shipping location. After decomposing the functions, the allocations of the functions can then be assigned to different actors. In the current example, these actors can be ground control operator, aircraft and lab technicians, as well as the aircraft itself. The model is intended to allow exploration of different allocation strategies and comparative evaluation with various team performance measures, such as number of information exchanges, physical resource transfers, taskload, busy time, and idle time.

IJtsma et al. [37] evaluated three allocation strategies, where joint activity is defined at the level of physical functions. The first had agents fixed to a general function and effectively all physical functions below it – no joint activity. This meant that if an agent was assigned a general function such as vehicle maintenance, only that agent and no other performed the physical functions below it, e.g., recharging batteries. The second allocation strategy grouped general functions to common physical or information resources. When the common resource was physical, the general function and all physical functions below it was then assigned to a single agent, provided the agent was also able to perform all the functions independently. The reasoning was that the sharing of physical resources would require coordination and communication; since the purpose of this model is practical and not social engagement, communication was considered overhead, and hence needed to be minimized. This meant that if an agent was assigned vehicle maintenance and there was only one battery charger, i.e., the physical resource, to perform the task of recharging the vehicle, only that agent was assigned to vehicle maintenance to avoid coordination overhead for sharing the charger. However, the model assumes that information resources are limitless, so agents can be jointly assigned functions that shared only information resources without communication overhead. For example, if a function was to determine if an aircraft had enough power for a mission based on battery capacity information, then that function can be jointly assigned to multiple agents because the battery information can be repeated to all agents simultaneously. With the charger, agents would have to take turns accessing it. In the third allocation strategy, the priority was to minimize the duration of a mission. This meant that general functions that were assigned in whole to a single agent in the second allocation strategy, can now be parsed out at the physical level to different agents. However, the physical functions that required a common physical resource stayed with one agent. The remaining physical functions can be assigned to different agents to be executed in parallel – independently, not jointly. This can occur even when not all parties have access to the same information. In the UAS example, this could mean that a vehicle can determine how many sorties to fly before returning to base from onboarding information about wind gusts that are dynamically changing the rate of battery consumption. Occurring in parallel, ground control operators could be operating cameras for a bridge inspection with information about camera position.

The performance metrics were aimed at comparing the function allocations based on various team qualities, such as efficiency and coordination, where coordination was considered overhead. Except where joint activity was expressly necessary, such as working together to lift a heavy object, the function allocations were structured to avoid interaction, and parallel activities were preferred to speed-up performance. It is also important to point out that, unlike the model presented by Pacaux-Lemoine et al. [129], IJtsma et al. [37] assumes that all information, including contextual data, is knowable, and can be explicitly obtained – a positivist view. Thus, it was considered better to exclude communication, a social interaction, from design of the system, not recognizing that there is information that is not readily knowable that the system needs, such as latent contextual information from humans about changes in intent and operational goals. Rather than avoid social interactions altogether, e.g., communication, contextual information can still be fully leveraged, but the interactions need be designed within both the limits of the machine, as well as the limits of the human so that it does not become overhead. I discuss this next.

The Human Autonomy System Oversight (HASO) model [131] draws on extensive empirical work to inform how the human is expected to respond to various automation interaction paradigms (e.g., supervisory control), and the quality of the automation (i.e., robustness and reliability). The relevance here is that HASO can be applied to identify which engagement strategies should be explored for addressing the data solicitation problem, based on the cognitive constraints that would impose on the human's ability to respond to prompts for providing data. The constraints are workload; engagement; and complexity. For this discussion, I define workload to be perceived cognitive effort as a function of the difficult and number of tasks to be accomplished [132]–[137]. Also, I limit the definition of engagement to be the duration and frequency of interaction with a system [15]. Complexity is the perceived complexity of the system that comes from, for example, the number of features and modes present in the system and how much data the system requires the human to comprehend. The HASO model predicts that as the level of automation increases from manual to fully autonomous, the perceived complexity of the system will increase, but while workload and engagement would decrease. Workload, engagement, and complexity do not interact, according to HASO, but do influence situation awareness. For the purpose of this discussion I loosely define situation awareness as the awareness of the system state and behaviors as well as events and relevant information in the operational environment. Situation awareness is important because it determines the human's effectiveness in recovering functions when or if a system fails or when a human is needed for decision-making in a supervisory interaction structure [138]. HASO was informed primarily by findings from LOA research. In considering joint activity against the constraints described in HASO, some initial questions can be considered. Would we continue to see a reduction in workload and subsequent reduction in SA as a result of implementing joint activity at the high end of the LOA spectrum? What impact will joint activity have on workload? If the coordination required by joint activity increases workload, what impact will this have on the available capacity for engagement and the opportunity to solicit data from users? Next, I review human machine teaming interaction design principles to address some of these questions.

Shively et al. [139] proposed human machine teaming interface design tenets. These tenets[12] are transparency, operator directed authority and bi-directional communication. Recall that transparency typically requires making information about the underlying reason for machine behavior available and understandable to people [110]. If there are machine agents in a team, and the behavior of the machines are impacted by their capacity, transparency may reveal the relationship between that capacity and the behavior. By doing this, it may also help to shape the perceived complexity of the system so that projections about the machine behavior can be more accurate, allowing other agents to coordinate their own actions effectively to move forward on a task. If the underlying reason for a behavior of an agent reflects contextual constraints, that information may be propagated through transparency to other agents in a team as well, so that other agents can be aware and take action to compensate for any gaps in performance with complimentary capacities. Shively et al. [139] argued that transparency is important because it affords humans the ability to align perceived capacity against the actual capacities of the machine. The result then is that actual performance can meet expectations, which in turn may afford positive trust calibration and improved engagement (see Chapter 5). Operator directed authority is a design principle which asserts that interfaces should allow dynamic allocation of functions between humans and machines while the human remains in control of when and how those allocations occur. In addition, the previous tenet emphasizes that agent independence is exercised as coordinated, but scripted action, e.g., recall the football playbook analogy. Such operator directed interfaces accommodate for the human constraints described in the HASO model [140], such as situation awareness, attention and workload. Operator directed authority is inconsistent with coactive design and mixed-initiative principles, both of which is an acknowledgement that automation is still brittle, and humans will be needed to provide direction and corrective inputs, at least for the foreseeable future. Finally, bi-directional communication bet20ween the human and

[12] This list is not a complete list of tenets proposed by Shively et al. [139]. For the complete list, see referenced paper.

machine supports shared situation awareness. This affords machine awareness of the state of the human, as a form of machine-of-human transparency [42], so that it knows when to engage and prompt the human for information. For the human, bi-directional communication supports machine transparency. Shively et al. [139] proposed that a human machine teaming agent be designed to act as an intermediary between the human, other machine agents and functionally specific automation. The agent would act a conversational recommender by assisting the human with different decision options, as well as make projections about the potential outcomes of each action. In addition, such an agent will track overall progress of reaching team goals, as well as issue direction to other machine agents on behalf of the human. Development of such agents for bi-directional communication provides an opportunity for testing various social engagement strategies for enhancing user experience, as well as create opportunities for obtaining information from users.

Tokadli et al. [141] has recently addressed the need for an easy to use interaction design look-up table. The table maps potential interaction paradigms, such as conversational interactions, haptic, physical (e.g., lifting heavy objects and retrieval), and gesture, to criteria required to accomplish a task. The criteria were derived from a combination of human machine teaming dimensions and work domain dimensions. The human machine teaming dimensions include type of system, such as robot or unembodied (e.g., recommender systems), the ratio of human to autonomous agents and their roles (e.g., supervisor or decision-maker), and team processes (e.g., coordination, cooperation, and communication). Work domain dimensions include context (e.g., outdoor, mobile, and stationary), task type (e.g., cognitive or physical), and conditions (e.g., nominal or emergency). One potential way to apply the table is to first identify the most effective teaming structure, perhaps from modeling efforts like those proposed by Pacaux-Limoine et al. [129] and IJtsma et al. [37] and then consider the work requirements to identify what combination of interaction paradigms to implement. Tokadli's [141] framework points to conversational interactions as an implementation solution. However, it does not provide guidance on how such interactions should be achieved. As discussed in Chapter 5, an implementation of conversation interaction will need to incorporate social engagement techniques, such as those employed in both visual and verbal anthropomorphic interfaces. These approaches, although may be effective in human-to-human interactions, have yet to be validated empirically for human-machine interaction. I discuss future work to address this gap in the next chapter.

In this chapter I provided a selective review of human machine teaming frameworks that describe interaction design principles motivated by joint activity. The Cognitive Control Model, coactive design, and mixed-initiative interaction frameworks provide high level principles for and priorities for exploiting autonomy. Pacaux-Limoine et al. [129] and IJtsma et al. [37] provide methods for evaluating and making decisions about what teaming structures should be applied. Endsley [140] identified relationships between the behavior of automation and cognitive processes to identify what human constraints need to be considered when implementing human machine teaming interactions. Shively et al. [139] and Tokadli et al. [141] offer some early notions of how to embody those interactions in interfaces, so that autonomous agents can be employed in the near-term. The work of Shively et al. [139] and Tokadli et al. [141] may be important in addressing the data solicitation problem. However, this remains untested. Empirical work will need to be conducted to determine any improvement on data solicitation with joint activity. In addition, interaction strategies will need to be evaluated using human factors criteria, such as situation awareness, workload, and trust and transparency, to identify potential negative impacts caused by data solicitation activities [140].

The principle claim of this thesis is that engagement approaches to development of user interfaces can be applied to improve the likelihood of obtaining data from users. In the next chapter, I conclude with a summary of the arguments made to support this claim based on a selective review of the literature. Subsequently, to stage a discussion about future empirical work, I discuss some potential application areas, i.e., self-driving cars and aviation, that might serve as an appropriate context for testing the concepts presented herein. In addition, I provide additional considerations for the formalization of social discourse for machine implementation and the social constructs involved in that formalization. Finally, I close with a research agenda for taking this research forward in future empirical work.

7. FUTURE DIRECTIONS

In this work I characterized the data solicitation problem and identified its importance to the recommender system application area. Subsequently, I suggested that engagement approaches leverage anthropomorphism to serve as a natural and unimposing alternative to brute force and tedious prompts for data. I presented a selective review of representative human automation interaction frameworks to capture current efforts in the area. The review showed that the research community is pivoting from human-centered frameworks to exploit machine autonomy for interaction structures that facilitate shared control authority between human and machine agents. These human machine teaming frameworks are only starting to take shape and aim specifically at interactions that facilitate practical joint activity for accomplishing work. These frameworks recognize that information exchange between all agents in a system is important to joint activity, but do not provide a structured approach to obtaining that information from humans. From the theoretical arguments made in Chapter 5, interface concepts for eliciting engagement may provide some structure, but their effectiveness in obtaining data remains largely untested for recommender systems. Although some work has been done in healthcare to evaluate the effectiveness of anthropomorphic interfaces for soliciting medical information from patients, there is little evidence in the literature that can inform or justify a design of such interfaces to obtain a broader spectrum of information, including contextual data, for recommender systems. In this chapter, I close with a discussion about interface concepts for engagement with machines that are based on formalizing social discourse, as well as discuss their limitations. I add an introduction to communication practices in air transportation Crew Resource Management to provide some additional insight into the development of such interface concepts. Finally, I suggest future work and present a framework that encapsulates all the constructs presented herein, along with the relationships among them. From this framework I point to prime research areas to hopefully guide any future work outside of the current effort.

Self-driving cars afford spare workload and attentional capacities because the human no longer carries the burden of manual control [38], [142]. Lee et al. [38] refers to this as the benefit of automated driving. This benefit suggests humans can direct the spare resources to productive activities such as work-related teleconferences, or in other business models, the driver engages the car interface for commercial activities [10]. However, it is unlikely this benefit will ever be available, as humans will typically engage in other distractions; this includes interacting with a phone, which is a common distraction even in manual operations [143]. A sobering example of this distraction, is the Uber self-driving car accident in Tempe Arizona where onboard cameras showed that the safety driver was eyes down on a cell phone for 7 of the 22 minutes of the trip and 5 seconds before striking and killing a pedestrian[13] [144], [145]. Unfortunately, more examples can be found with Tesla's autopilot feature, where drivers have been killed because they failed to re-take control when the vehicles collided with cargo trailers it did not detect [144]–[147]. Lee et al. [38] attributed the failure to re-take control of the vehicle to what they called an inability to recover the cost in driver readiness that was lost to not being engaged in manual driving. Referring to the HASO model [140], SA can be an index for readiness, which diminishes with higher levels of automation [148]. One means of mitigating this tradeoff is through employing transparent interfaces; this is what Wickens [149] calls "a free lunch" referring to the benefit of exploiting higher levels of automation, but without the cost in SA that comes with it. For self-driving cars this can mean that the cost in driver readiness can be absorbed if the vehicle interface is providing transparency to the user through an ongoing exchange of information. As discussed in Chapter 4, this transparency can emerge from information exchanges in both practical and recreational activity. There is currently no evidence in the corpora that suggests engagement has a cost or that it is cost free, even if it is enjoyable. Contrary to Wickens, then, depending on how transparency is implemented, it may impose its own cost in terms of workload and attentional resources. One can develop a solution for absorbing the cost in driver readiness, but in exchange for SA these payments in cognitive resources may still need to be made upfront and continually throughout the duration of the operation. This way, one has the information needed in hand to stay ahead of the time horizon for recovering a vehicle. In the end, the adage, "there is no free lunch", unfortunately, may still hold. Presuming there will be a cost, the ceiling should be no higher than manual control. Based on the above arguments**,** I propose that a conversational interface that facilitates engagement

[13] The NTSB report also noted that the company had disabled safety features on the vehicle that could have prevented the accident. These were the obstacle alerting system and automatic braking originally enabled in the production vehicle from Volvo. Uber expected the safety driver to be responsible for recovering the functions originally provided by the disabled safety features.

should realign those resources as part of joint operation of the vehicle so that they are not “wasted” on distracting tasks.

Engaging interactions can be employed to displace distracting secondary tasks to re-engage the driver with the operation of the vehicle. It can serve as what Baldwin and McCandliss [142] refer to as an incentive to maintain vigilance. As demonstrated with video games, attention can be held exceedingly long, if the task is enjoyable or interesting [150], while the perceived workload can still be present. If engaging interfaces were designed to offer such incentives it might also provide similar gains, but the interfaces can be directed towards accomplishing work, as opposed to gaming. For the Tesla Autopilot, a conversational interface approach can be designed to entice the human into joint operation of the vehicle. This joint operation does not necessarily mean that the human has to return to manual operations, but it can mean assisting the system by offering contextual information the way humans would do for one another [127], “hey, watch out for that trailer, he looks like he is about to turn.” Such a design would imply a shared control model more representative of a copilot[14], rather than an autopilot, a term that carries the misleading notion that the human can be detached from the vehicle operation[15]. In the Tesla Autopilot example and the Mars Rover example from Fong’s Collaborative Control Model, engagement can be employed to facilitate joint operation of the vehicle through an exchange of contextual information. Based on these examples and the theoretical arguments made herein, it seems reasonable that engagement can elicit more information from people, but this relationship needs to be verified empirically. Furthermore, any empirical work will require operationalization of engagement to structure the interaction with users, like the anthropomorphic verbal and visual presentation techniques described in Chapter 5. When employing engagement to obtain information, one also needs to be careful not to mislead by inaccurately implying that the machine possesses human capabilities associated with looking and sounding human. The ramifications of this is that humans will quickly disengage as their tolerance for machine error is very low, when compared to other humans. For example, when Bob hears a call to pick-up a cappuccino for Rob at a café, he may come forward

[14] A copilot system implies that the human still carries principle responsibility for the operation of the vehicle, thus addressing automation complacency issues associated with the term Autopilot.

[15] Elon Musk, the CEO of Tesla, has defended the use of the term autopilot by stressing that in aviation pilots are expected to actively manage the aircraft even when autopilot is on. However, by making this point he ignores the fact that he is marketing to everyday consumers, not seasoned commercial pilots, or instrument rated pilots to the least.

to verify the order was for him, rationalizing that the order was misnamed due to a listening error caused by the environment – loud music and chatter. If a machine made the same error, it may be attributed instead to a defect in the machine; people would disengage, presuming the error will repeat due to the inherent flaw. To overcome this, Wallis [151] suggests a strategy for keeping users "in the flow" of a conversation so that they overlook these mistakes as they would in natural conversations with another human. I discuss this further next.

Wallis [151] asserts that if the initial agenda is not to test the reliability of the machine then most people will go along with the conversation, even when the machine is factually incorrect. We do this commonly in conversations between people, particularly when the facts are not immediately verifiable, "Those talons are so big they can pick up a bear." In this example, it is factually incorrect, unless this was a lecture in paleontology, to state that there are talons that can pick up a bear. More importantly, this example shows that we often model the intent of the speaker to make projections about what is being communicated. Intent is the latent contextual information residing inside the mind of a human that describes the personal goals of communication and action. Intent is modeled because people do not always explicitly convey their intent to others. With an accurate model of intent, one can project what someone is trying to achieve through behavior and speech. In this case the intent may be that the speaker wants to convince the listener of the size of the talons on a bird of prey.

Wallis [151] suggests that conveying intent serves to fuel conversation. This allows people to determine if others share similar personal goals. If the goals are the same, one would expect very little conversation to occur because there is nothing to negotiate. When people have competing goals, they engage in what he calls dialog games, where the participants have goals and plans for achieving them. The objective would have to be more concrete than arguing about the size of a bird's talons. I was in a bazaar in Bali looking for a fruit stand and stopped at a clothing store to ask for directions. In this scenario there are conflicting interests; mine was initially to buy fruits, while in exchange for information, the shopkeeper urged me to listen to a pitch for some batik clothing and have me walk away with some merchandise in tow. The example here is intentional conversation. In intentional conversation, speakers in the dyad will model each other's intent and fill-in for inaccuracies to pursue their personal goals[16]. If the shopkeeper indicates a size medium shirt while not noticing the label showed a size large, the conversation would not abruptly stop. I would fill-in, perhaps by rationalizing that he just wants to advertise the design first and work out the size later if I commit. For the sake of keeping with the flow of the conversation, and for me, quickly getting to the fruit stand, I looked past the inaccuracy of the shirt size. Similarly, Wallis suggests that if a machine also conveyed goals, people would engage in continued discourse with the machine to resolve competing goals the ways humans do in the example above.

When language is used to manipulate others to achieve personal goals we have what Wallis referred to as the dialog game [151]. The outcomes of the game are described in a goal space; this is structured as shown in Figure 9. What Wallis wanted to do was describe the results of the dialog game between a human and machine. In region A, neither the machine nor the human has intent, this is the region of gossip and casual conversation. Unlike between people, designing a strategy based on casual conversation will yield about four minutes of engagement before the human abandons the machine[17] due to boredom [152]. Wallis recommends that a machine introduces an intent if initial conversations do not yield an intent from the human within the 4-minute time frame. When either the human or machine brings intent into the conversation, they move the dialog game to one of the remaining regions. Region B is the only region where the machine may win because the human has no intent and defaults to the goals of the machine. Alternatively, in region F, the human can simply disagree with the intent of the machine and refuse to discuss it by disengaging. Unlike the shopkeeper in my personal example, it is easier for a human to walk away from a machine than it is to walk away from another person. The distinction here is that the machine is assumed to be socially inferior to the human [153], so where the goals of the human and machine[18] are in conflict, the machine must always yield – there are more ways for the human to win. In region C, this is straightforward; the machine has no intent and yields to the human's intent. Region E represents a machine that can presumably refuse to accept the human's intent and disengage, but

[16] This refers to everyday conversations, not formal academic debates, where the arguments are judged on the facts used to support them. Although, some academics insist on this approach to everyday life to only find themselves isolated to a special community.
[17] We are willing to exceed 4 minutes to chat with other humans because there is social reward for doing so. With machines this social reward, e.g. verbal praise, may be less convincing, especially if there is no similar social history.
[18] The machine has intent to the extent it reflects the goals of the designer.

because of the social distance, the machine typically does not have region E to its favor. Similarly, in region D, both the machine and human have intent, but then yields to the human. In the event that a machine refuses the human's goal, perhaps due to a fixed constraint, like a self-driving car that will not proceed, despite continued input by the human, because it detects an approaching pedestrian that the human has not seen, then the machine should explain why – provide transparency.

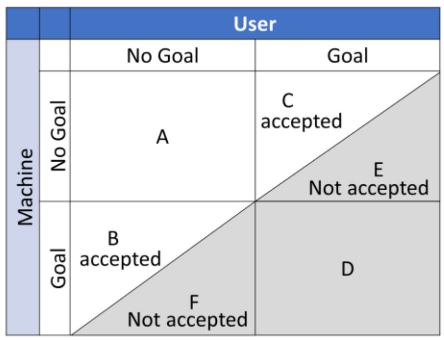


Figure 9. The goal space for a two-agent conversation. Adapted from Wallis [147].

The social discrepancy between the human and machine makes it difficult for machine agents to carry a conversation because the machine does not have the agency to pursue their intent. Recommender systems lie in region B and F[19]. The extent to which RSs engage in conversation is limited to the discretion of the human; they can prompt to bring attention to solutions, but it cannot press a solution like humans can if they thought it was important. In the real world, when a machine's intent is rejected or ignored there is no recourse - unless the machine serves to represent a human or organizational authority. We are reminded of this each time the university's bursar system prompts us for tuition, or we get caught in a phishing scheme. What if in a mixed-initiative interaction design the machine is given more authority to assert engagement with a human? Would it result in more dialog and afford more information from people? The discrepancy in social ranking that we see between humans and machines is not unique to that dyad, it is just that in a ranking among people, machines will occupy the lowest part of the totem [153]. Would it be

[19] The human here still has intent, and that is to find a solution to a problem. However, they are relying on the RS to generate solutions that they do not have on hand. The RS displays intent through its recommendations and succeeds if the human accepts one of them. Otherwise, it fails if the human disengages and takes none of the options.

effective then to elevate the authority of the machine so that it is seen as a team player, as has been proposed in human machine teaming frameworks? How would the interaction be structured? The work by Cassell and Bickmore [104] provides some insights to these questions.

In the concept developed by Cassell and Bickmore [104], the agent is designed to autonomously interleave casual conversation with task oriented conversation. Referring back to Wallis' [151] goal space matrix (Figure 9), this is similar to moving between region A and then other cells where either the machine or the human have goals. The idea is that the casual conversation, termed small talk, is used to catalyze both goal-directed conversations and continued small talk.

"Hey hot day today, kind of brings me back to Hawaii, but without all the ocean and sand. About this weekend, I kind of want to postpone the cabin reservation to next Saturday."

In the above example, a speaker initiates with a potentially interesting topic about Hawaii, the small talk, and then transitions into a proposition to reschedule a cabin reservation, the entry into goal directed discourse. To keep the discourse natural, the topics needed to be context sensitive and build on information learned from previous exchanges. This was achieved by implementing a discourse planner. In this planner, the agent guided the user through different topics, which were represented as nodes in an activation network[20] [154]. The topics in this implementation were predefined[21], and the dialog was taken from actual conversation between humans. Transitions between different topics, or nodes, were based on the path that maximized information gathering, maintained coherency in the conversation, were consistent with user preferences, and minimized face threat. Information gathering was maximized by prioritizing transitions to conversation topics that would query or lead to information about the user's preferences, "I see that there are a few homes on the market right now that you might be interested in, but before we go into that, I'd like to know a little about your lifestyle." Nodes that maintained coherency in the conversation followed-up on topics that were previously discussed, so this system would avoid nodes that involved topics about cats if the current topic was about birds. The conversational agent was

[20] The activation network was originally designed to enable action planning for autonomous agents. The nodes represented an action a machine can take. Transitions between nodes were based on whether goals and capabilities that were defined in the next node met the demands imposed by what the machine knew about the context.

[21] The researchers were building a virtual real estate agent, REA. Users were given the expectation that topics would be limited to REA's expertise. However, the agent was able to propose unrelated content during small talk, like weather. When on a task related topic, REA asked about preferences on a home, such as number of bedrooms.

sensitive to what it learned about user preferences; if the user expressed an interest in birds then transitions to nodes about birds were preferred over nodes about cats. Maintaining coherency in the conversation kept the agent on topic about birds once it was started. According to Goffman [155], face is how people want to be viewed. Politicians and celebrities are perfect examples of people who constantly manage their public image to minimize face threats. In Cassell and Bickmore [104], face was treated as a predictor of whether someone would want to engage in a conversation and determine what small talk topics to pursue. When that face is threatened, it is motivation for avoiding discourse. Face threat is influenced by a predefined set of factors based on known social constructs. I discuss these social constructs next.

The constructs considered in the face threat model were power, solidarity, familiarity, affect, and the intrinsic threat from certain types of speech [102]. Power is the ability of one person to control the behavior of someone else. Face threat diminishes when this is false. Solidarity is the like-mindedness between people. Individuals who share a profession, political beliefs, religion, and values are said to be like-minded. This was hardcoded into the agent but learned in active conversation with the user. If a topic expressed likeness with the user, then that diminished the face threat. Familiarity is the extent to which there is a reciprocal exchange of information. This was quantified by the number of topics discussed between the agent and user; the greater the number of topics discussed, the lower the face threat. Familiarity also possessed depth; this was quality of the information that was either public or private. Requests for public information reduced face threat, while requests for private information increase face threat. Affect is the degree to which interactants like each other. If people like each other, then face threat is diminished. This would be determined as the conversation progressed with the user; the more conversations the agent has with the user, the better the affect prediction. The intrinsic threat from certain types of speech comes from the perceived intrusiveness of the speech. Speech that is intended to inform are ranked less threatening than requests for information. Speech that involves rejecting someone's ideas is ranked most threatening. Although Cassell and Bickmore [104] addressed a wider number of social constructs to achieve a face threat model that was as complete as possible, for scoping the problem, they only applied familiarity and solidarity, among the five, to choosing a small talk topic during a conversation. In addition to those social constructs, small talk topics were limited to what was learned about a user's preferences and had to be related to the topic on hand.

Cassell and Bickmore [104] tested their concept by having users interact with an animated speaking agent. The agent posed as a real estate agent and always initiated the conversation with users, thus, showing agency. They had two experimental conditions. The first exposed users to a discourse strategy that involved only task-oriented speech. In the second condition, small talk was interweaved with task-oriented topics. Their evaluation included user ratings for trust, whether users thought the agent knew about them and their preferences, whether users thought they knew the agent (i.e., can be considered a measure of transparency), whether users thought the interactions seemed natural, and whether the agent was engaging. The relevance of this evaluation on the current work, is that positive outcomes for these measures could, presumably, indicate that continued discourse is achieved. This, in turn, creates more opportunities to obtain data. In the analysis, the researchers crossed the experimental conditions with two different characteristics of users; introverted versus extroverted, as well as passive or initiated. The passive and initiated classifications were made after they discovered that some users, ignoring that the agent was designed to initiate the conversation, attempted to control the dialogue by initiating themselves. The passive users simply waited for the agent to initiate. It is also important to note that the researchers did not find any correlation between introversion and extroversion with passive and initiated. The results were as follows.

Introverted users revealed no difference in ratings of trust between the small talk condition and the task-oriented condition, while the extroverted users expressed greater trust in the small talk condition. When asked if users thought the agent knew them and their needs, extroverts gave higher ratings in the small talk condition. Results for the extroverts were inverted for this measure, giving higher ratings in the task-oriented conditions. When asked if users thought the interactions were natural, extroverts gave higher ratings in the small talk condition, while results for introverts were again inverse. The analysis for engagement and whether users thought they knew the agent (i.e., the transparency measure) were crossed with passive versus initiated. There were interaction effects for both. Initiated users gave higher ratings for engagement in the small talk condition, while passive users gave higher ratings in the task-oriented conditions. Similarly, on the measure of transparency, initiated users gave higher ratings in the small talk condition than in the task-oriented condition, while results were inverted for passive users. This result is somewhat unintuitive because one would expect that if a system is effectively providing transparency then it should be universally true across all users.

Findings from Cassell and Bickmore [104] emphasize the importance of having an updatable and accurate model of the user during social discourse. As shown, user characteristics dramatically change their response to the conversational strategy. Requiring a system to form a user model who is new to the system does reintroduce the data solicitation problem. However, it appears from these findings that if machines are afforded the agency to be proactive in engaging the human, as suggested by human machine teaming principles, the data solicitation problem can be effectively addressed. Even when some users are not amenable to social interaction, such as introverts, a proactive machine can initiate to at least learn if a more conservative conversational strategy should be used. Whether or not a fully animated conversational agent such as the one developed by Cassell and Bickmore can solicit more data from users for recommender systems remains to be determined. To address this, I conducted an exploratory study where a similar agent was employed.

The exploratory study (Appendix A) aimed to investigate if user interaction with a full-featured anthropomorphic agent, like the one presented in Cassell and Bickmore [104], would positively affect the number of ratings that could be obtained for recommender systems. The idea was that a machine agent would leverage the social discourse generated by visual and verbal cues to coax a user into responding to more prompts for ratings data. Ratings data was selected for this study because it is currently the most important to developers of recommender systems.

A 2 x 2 within-subjects design was employed (Table 1). Anthropomorphism (levels: no agent versus agent) was crossed with Interaction Method (levels: mouse versus voice). In the no agent condition (Figure 10) a synthesized voice prompts for ratings was played over a speaker, e.g., “What is your rating for the cinematography?”, along with a display of an object on a monitor. The agent condition (Figure 11) provided the full-featured anthropomorphic presentation. The anthropomorphic presentations had a fictitious back story and the voice synthesis mimicked natural speech that included the use of pronouns to refer to the system as if it were a real person. Unlike in Cassell and Bickmore [104], the speech was fully scripted, but the study was designed so that a single participant never got the same script twice. Interaction Method was included in this design to examine possible interaction with Anthropomorphism. In the mouse condition participants reported ratings by selecting the desired number of stars with a mouse. In the voice condition participants reported ratings using voice recognition through a microphone. One might expect greater engagement, and more ratings, when voice recognition is paired with agent presentations because this is consistent with human-to-human interaction. However, when the input method is inconsistent, e.g., mouse with agent presentations, one may expect the inconsistency to suppress engagement and reduce the quantity of ratings.

Table 1. Experiment 1: 2 x 2 within-factors Anthropomorphism by Interaction Method.

Anthropomorphism

Interaction Method

	No Agent	Agent
Mouse		
Voice		

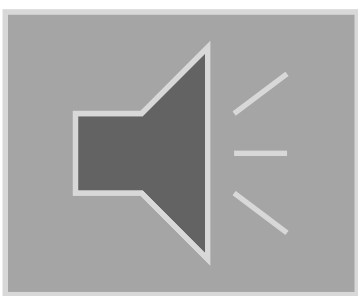


Figure 10. No agent static object.

Figure 11. Agent.

I offered the following hypothesis.

H1. The quantity of ratings will be the highest when agent presentations are paired with voice recognition inputs.

H2. The quantity of ratings will be lowest when no agent presentations are paired with mouse inputs.

The experiment resulted in no main findings, so no conclusions can be drawn on whether anthropomorphic presentations can elicit more data from users. In general, participants gave all the ratings requested regardless of the condition. Potential reasons for this was that either the task was too easy, or participants felt obligated to provide ratings as part of contributing to the study. That is, users gave the ratings because they felt it was part of the instruction, and at only 10 requests, this may not have been difficult to do so. In a recent study [156], researchers found that participants typically give-up on providing ratings at about 17 and 18 requests. Thus, a follow-up study would need to employ up to 20 and 25 data requests to possibly see any changes in quantity of ratings from users. As an exploratory study, it is important to note that the conversation between the human and machine was not dynamic and did not go beyond simply providing or rejecting requests for ratings as the task. Unlike in the agent implemented in Cassell and Bickmore [104] the exchange was very simple with an emphasis on visual presentation. This approach was taken because the study aimed to first establish any general impact of anthropomorphism. Generating convincing social agents requires a considerable amount of resources. It then behooves the researcher to determine if any general effect of anthropomorphism can be found before further investment in developing machine agents. In addition, this study targeted ratings data; it is still unknown what impact the manipulation here would have had on obtaining contextual data. Finally, although there were no main effects, the data collected from questionnaires did show that participants found the agent anthropomorphic and engaging. Therefore, the presentations concepts created here can be used to manipulate anthropomorphism in future studies.

In addition to controlled studies, future work can examine interaction strategies in various application areas to determine the resilience of the solutions. The self-driving car application was one that was considered earlier. However, for early implementations a more structured environment with very narrow topic focus may be a more convenient place to start. If for example, an activation planner based on a network of topics was being implemented, such as in Cassell and Bickmore [104], a narrow topic area would reduce the number of nodes in the network and be computationally less taxing. Also, it would make transitions between topic nodes more predictable. Another important contribution that Cassell and Bickmore made was they demonstrated how to incorporate complex social constructs into a working computational model for engaging people. These social constructs, such as power and solidarity carry considerable relevance across many applications areas. One such area that can serve as an appropriate platform for future work, because of its rigid and predictable structure as well as for its relevance to the social constructs that shape machine agent conversational strategy is air transportation. In air transportation there has been a history of social discrepancy between members of the flight crew, which has led to poor communication and poor performance. Major improvements in the flight crew's social interaction has been heavily influenced by a discipline called Crew Resource Management (CRM).

Although having broader application in any area that involves teamwork, such as offshore oil production [157] and surgery [158], CRM had its genesis in aviation, so the definition I apply here refers to safe and efficient flight operations by exploiting all available resources; resources that include information, equipment, and people [159]. CRM encompasses training, team structure, decision-making, culture, and communication [160]. Entire textbooks have been dedicated to the topic, so a complete discussion is out of scope for here. I will focus instead on communication aspects of CRM to describe conventions for information exchange within flight crews, because it is the most relevant to the data solicitation problem.

CRM communication can be verbal or visual, e.g., a copilot pointing to the altitude window on a flight control unit and announcing the target altitude after making a change to the assigned altitude. Similar to Clark's [43] view on language and joint activity, CRM communication is not limited to simple information transfer, and must include functions pertinent to team activity; these are to establish team relationships, establish predictable behavior and expectations, maintain attention to task for shaping situation awareness, and to serve as a tool for managing time and taskload. The ability to achieve these functions determine the effectiveness of flight crews in solving problems when nominal or off-nominal events occurred; that effectiveness was mostly contingent upon the pattern of communication adopted by the crew members. Among these patterns was engaging in frequent interchanges during low-workload periods [161]. These interchanges include recognizing problems, stating goals and subgoals, planning and strategy formation, gathering information, and alerting and predicting. During these interchanges effective flight crews structured the communication to proactively express intent to perform actions and closed the communication loop with frequent acknowledgments [162]. In modern flight decks, where an autopilot is being implemented, i.e. as opposed to the Tesla feature that adopts the reference, pilots are professionally trained to remain fully engaged in the operation of the vehicle without the need to continuously provide inputs into the controls. I suggest here that a similar interaction strategy can be used in human machine teams for autonomous vehicle operations, providing what Wickens [149] called the "free lunch" through transparency, and closing what Lee [38] referred to as the cost in driver readiness. CRM communication strategies is a potential solution to designing joint activity and socially engaging the human to achieve better data solicitation as a seamless part of making meaningful contributions to the operation of a vehicle. In the commercial flight deck, CRM is vehicle operation, not a substitute for manual control. As implied above, this means that pilot training has incorporated tasks that pilots need to do when they are both manually flying the airplane and when the automation is controlling the aircraft. These tasks involve aviating, navigation, and communication. All three of these responsibilities entail operation of the aircraft and workload is shifted within and between them. Unlike in self-driving cars, CRM training has been designed to minimize the "waste" of spare capacity to distractions.

To conclude, I now summarize the intended contribution of this work. That contribution is a framework to guide scientific investigations into interface design concepts that will address the data solicitation problem (Figure 12). The constructs covered in this work can be separated into two categories, interface design concepts and outcomes. As an outcome, **user data** comes in terms of quantity and quality. The expected impact of user data is that it will improve overall system recommendations for users. When the system has less user data to ingest, the recommendations are poor because they become less relevant to the user. Although also an outcome, **engagement** has the role of keeping the user in the interaction loop. When interaction experiences with the system are positive to the user, they are likely to return to the system for further interactions. The type of interactions that appear to generate the most information exchange are those that involve joint activity. **Joint activity** is an interaction design approach that involves designing a machine to serve as an agent in human machine team structures that allow the machine to take on more proactive roles by, for example, reaching out to the human for assistance or offering assistance without it being requested. The presumed relationship between engagement and joint activity is that engagement can entice a user into joint activity with a machine where it is expected that the ongoing exchange of information will generate more user data. Both the relationship between engagement and joint activity, and the expectation that joint activity will generate more user data, are hypothesized and will need to be validated in future empirical work.

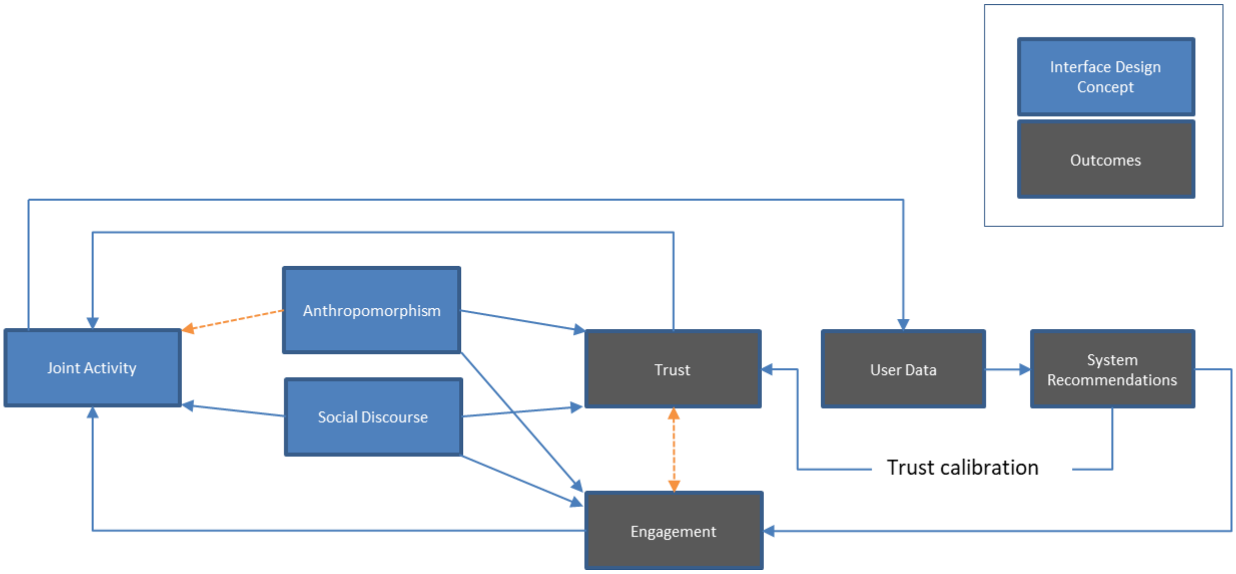


Figure 12. A conceptual framework on the direct and indirect influences of interface implementation concepts on engagement and user data output. All arches represent existing relationships identified in the empirical work. Dotted arches represent relationships not well supported by data and are prime research areas.

The **anthropomorphism** and **social discourse** interface design concepts come with more evidence to support their relationship to joint activity and engagement. These design concepts are complex and difficult to implement; empirical investigations will need to be conducted to determine what elements of anthropomorphism and social discourse are positively impacting engagement and joint activity. Anthropomorphism may lend initial trust to the system because of the human-like qualities and can lead to mis-calibration of that trust. This initial trust can be adjusted as system performance is revealed through repeated exposure. Findings from recent studies show a positive relationship between anthropomorphism and continued interaction with a system, but it remains unclear how that relationship can change if the anthropomorphism is negative, such as the uncanny valley [163]. Thus, the relationship between joint activity and anthropomorphism is a prime area of research. Finally, trust is presumed to have a positive impact on joint activity and engagement. Although there is growing interest in the research community for evaluating trust against novel interaction concepts like human machine teaming and associated joint activity, as well as engagement, the relationships that are being drawn between these constructs are mostly hypothesized, with evidence validating the relationships being scarce. The relationship between trust and engagement is an area of prime research interest because the direction of the relationship is ill-defined; but these constructs may have an impact on the overall quality of the interaction experience, and as a result have an indirect impact on obtaining user data. The research, e.g., [164], typically suggests that engagement influences trust, however, this could be an artifact of the experiment design. These studies are generally designed to determine the quality of an interface and trust is presumed to be an indication of engagement. The constructs of trust and engagement have not been distinguished, so future work should be directed to address this as an initial step.

I propose that future work to address the data solicitation problem for RSs should be as follows. The first is to test two key assumptions made in this thesis:

- Joint activity will lead to more data from users
- Obtaining contextual information leads to more relevant RS recommendations

Presuming that a system can obtain more contextual data from users, it has not been established that this type of data will lead to better RS recommendations. Although Adomavicius et al. [29] reported findings that users typically found results from context-aware recommender systems to be more acceptable than with results from ratings-based collaborative filtering alone, other factors, such as the quality of decision-making, as well as trust in the system, need to be considered. It is still unclear whether human teaming frameworks built on joint activity will lead to continued discourse and subsequently more information obtained from users. The current frameworks focus squarely on task-oriented activity. Generally, task-oriented discourse is designed to reduce coordination overhead and quickly close on a task. However, as was seen in Cassell and Bickmore [104] task-oriented discourse receive the best response from users who are introverted. For the extroverts, casual conversation helps to catalyze task-oriented conversations. Although Cassell and Bickmore did not tease this apart in their experiment, it appeared that to facilitate both social and practical engagement, a machine agent needed to convey goals as the conversation unfolded.

Future work should address the following for developing engaging recommender system interfaces:

- <u>Determine the impact of different aspects of anthropomorphism on joint activity</u>: What aspects of anthropomorphism will need to be determined. Is a full featured anthropomorphic experience required or just verbal interaction alone?

- Determine the relationship between trust and engagement, if any: The constructs of trust and engagement need to be teased apart, and any relationship between the constructs need to be examined.
- Model human social interaction: social constructs that apply to development of conversation planning must be determined.
- Model the human interlocutor: factors and mechanisms for modeling the users in conversation must be determined.
- Evaluate user reaction to novel machine roles: users' response to machines exercising greater autonomy in initiating interactions must be determined.
- Identify appropriate context for investigation: the best application areas for future work must be determined, along with advantages and disadvantages of each based on.
    - What applications afford the best opportunity for social interactions?
    - What applications will require recommender systems to be effective?
    - What applications will require information from the human in order to be effective?

As discussed previously, anthropomorphism can provide some immediate benefits in terms of improving engagement with recommender systems by leveraging existing tendencies for people to socialize technology. However, what approaches to anthropomorphic experiences are most conducive to obtaining data through joint activity needs to be determined. Trust and engagement play important roles in encouraging joint activity, but it is unclear if these constructs work together or independently to influence joint activity. Thus, studying the relationship between trust and engagement is a prime researcher area. In implementing a conversational agent for engaging people, social constructs needed to be incorporated into formalized discourse planning for machines. These constructs can be difficult to quantify; if part of the social discourse involves non-verbal communication cues with facial expressions and hand gestures, rendering the visual presentation so that it looks natural will be challenging. Although it is ideal to incorporate as many relevant constructs, not all of them are needed to make an effective conversational system for obtaining data from users. Future work should also aim to determine what is the minimal requirement on parameters e.g., interests, values and dynamics for developing rapport, that constitute a model of a human conversational partner. How humans will respond to a machine that displays these features to proactively pursue social discourse with them remains unclear, and deserves further investigation, as well. Finally, implementation concepts should be tested across different application areas. Air transportation and self-driving cars provide a convenient platform for examining the use of various conversation strategies to obtain data from users. They are ideal because the topics that can be entertained can be limited in scope and the rigid operational environment makes the social discourse with users predictable. However, operational demands in these systems can impose various human factors constraints on how a conversation system is implemented and how much information can be obtained from the human. In vehicle operation conversational strategies cannot overly burden the user with additional workload. The theoretical frameworks discussed herein would suggest that if the conversational system is designed to interweave data solicitation as part of vehicle operation there should not be any additional workload and interacting with the system would not be perceived as an additional burden on the user. However, there is no data in the corpora to support that conclusion. Any collateral benefits from engaging the user for data, such as improved SA through system transparency has yet to be demonstrated, as well. Positive results on trust scales from past studies when users interacted with a conversational agent shows some promise with respect to supported SA, but the relationship between trust and transparency will need to be validated before we can attribute those benefits to the user interface concept.

APPENDIX A. EXPLORATORY STUDY

Modeling human visual and verbal communication to solicit preference data brings with it both advantages and disadvantages. One advantage is that it can create engaging interfaces. With engagement designers can potentially coax users to remain in the interaction loop with recommender systems, and as a result, create more opportunities to collect ratings data. However, it is also possible that users are reticent about exposing their preferences and in the presence of a face, although artificial, would not provide ratings data. In Experiment 1, participants were asked to view full-featured anthropomorphic presentations that possess both visual and verbal aspects of human communication. This included a fully animated face, realistic skin texture, and natural voice synthesis. After viewing the presentations, they will be asked to provide ratings and evaluate the anthropomorphic quality of the presentations in a questionnaire.

Design and Hypothesis

The purpose of this experiment was to determine what effect, if any; a full-featured anthropomorphic presentation that modeled both visual and verbal human communication may have on the quantity of ratings. A 2 x 2 within-subjects design will be employed. Anthropomorphism (levels: no agent versus agent) was crossed with Interaction Method (levels: mouse versus voice). In the no agent condition a synthesized voice prompt for ratings was played over a speaker, e.g., “What is your rating for the cinematography?”, along with a display of an object on a monitor. The agent condition provided the full-featured anthropomorphic presentation. The anthropomorphic presentations had a fictitious back story and the voice synthesis mimicked natural speech that included the use of pronouns to refer to the system as if it were a real person. Interaction Method is included in this design to examine possible interaction with Anthropomorphism. In the mouse condition participants reported ratings by selecting the desired number of stars with a mouse. In the voice condition participants reported ratings using voice recognition through a microphone. One might expect greater engagement, and more ratings, when voice recognition is paired with agent presentations because this is consistent with human-to-human interaction. However, when the input method is inconsistent, e.g., mouse with agent presentations, one may expect the inconsistency to suppress engagement and reduce the quantity of ratings.

Table A.1. Experiment 1: 2 x 2 within-factors Anthropomorphism by Interaction Method.

Anthropomorphism

Interaction Method

	No Agent	Agent
Mouse		
Voice		

I offer the following hypotheses.

H1. The quantity of ratings will be the highest when agent presentations are paired with voice recognition inputs.

H2. The quantity of ratings will be lowest when no agent presentations are paired with mouse inputs.

Participants

Twenty-four participants were recruited from the San Jose State University Psychology Department and Purdue University because these institutions have a convenient and available mechanism for recruiting participants. The sample size was computed using G*Power [160], [161]. Course credit was given as compensation for participating in the study.

Procedure

Each session took approximately 90 minutes and included an introduction, information about the task, test trials, and a debriefing segment. In each trial participants viewed 3-minute movie trailers from three different genres: science fiction, musical, biography. In the agent conditions these trailers were preceded by an animated agent that presents a backstory and invited the participant to view the video which followed automatically (Figure A.1). When the end of the trailer was reached, participants were prompted by the animated agent for ratings in the agent condition. In the no agent condition a static object (Figure A.2) in the shape of a computer speaker appears on the screen and a prompt for ratings will be delivered with robotic speech. In the mouse condition, participants will responded to ratings prompts by clicking on a YES or NO link.

Figure A.1. Agent.

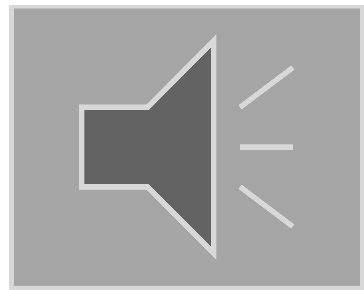


Figure A.2. No agent static object.

In the voice input condition, participants spoke into a microphone in natural language to indicate that they would either volunteer the ratings or refuse. If participants agreed to give a rating, then the display advanced to the next slide where they provided the ratings. If they declined, the display advanced to the last slide where the trial ended. In the mouse input condition, participants clicked "Yes" or "No" to respond to a prompt for ratings, followed by a separate slide where participants clicked on the stars that corresponded with the rating they wanted to give; Figure A.3 shows this on Slide 3 and 4 respectively for the agent condition. In the no agent condition the human agent was replaced by the static object in Figure A.2 and there was no backstory at the beginning of the trial. Dialogue for each ratings prompt from the human agent was unique. Examples including one for a backstory was given below each slide in Figure A.3. A complete list of the prompts and backstories presented in this study are in Appendix I.

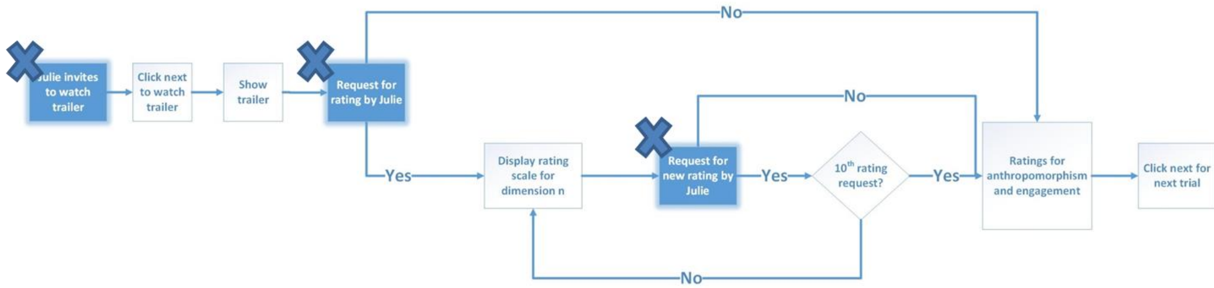


Figure A.3. Typical trial for the high anthropomorphism condition with mouse input.

In the voice input condition, the "Yes" and "No" buttons and clickable stars were absent, requiring participants to use speech for responding to the prompts. All participants were informed that they were under no obligation to provide a rating, and that there was no correct answer. A session ended when all trials were viewed.

Twelve unique movie trailers, each about 3 minutes in duration, were selected from YouTube. The trailers were distributed evenly across the three genres: science fiction, musical, and biography. The order of the trailers was counterbalanced across participants and each order was randomly assigned to a participant. A questionnaire was administered after each block of trials for a condition. Thus, questionnaire responses were multiplied by 4. Each session ended with a questionnaire that asked participants to compare the interaction methods against each other, and a debrief to acquire feedback from the participants about the anthropomorphic presentations (Figure A.4). The reason for the debrief is to provide context to measures described in the following section.

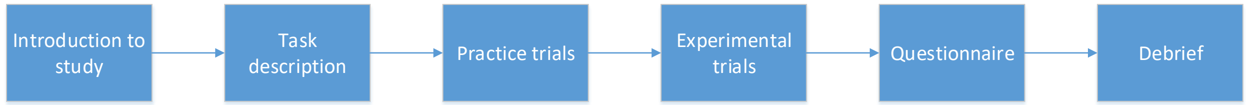
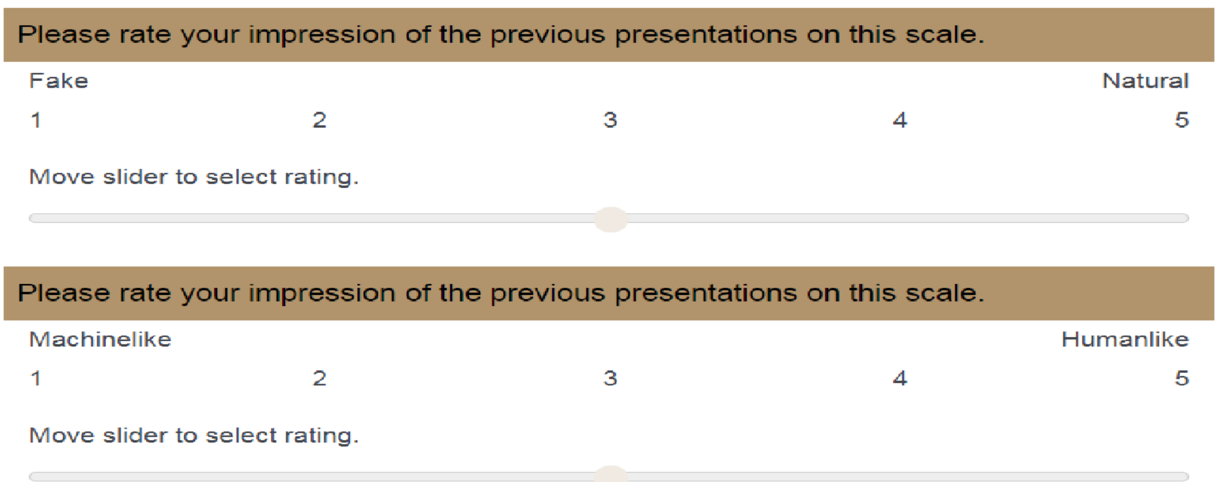


Figure A.4. Order of events for a single session.

Measures

The measure of primary interest in this experiment was to quantity of ratings. Each trial provided up to 10 opportunities to provide a rating, with a total of up to 120 ratings across 12 trials. Secondary measures were subjective ratings for the quality of the anthropomorphic presentation and preference for either mouse or voice input. The intended purpose of the anthropomorphism ratings was to validate the anthropomorphism conditions; subjective ratings for the high anthropomorphism condition should be higher than the in the low. The anthropomorphism rating scale was adopted from the Godspeed questionnaire created by Bartneck, Kulic, Croft, and Zoghbi [165]. Items from the Godspeed questionnaire were distilled from a literature review that identified key concepts related to engagement with robots; these are anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Each concept carried a number of dimensions that are believed to influence the concept. For example, anthropomorphism is believed to be determined by how real a robotic agent appeared, or how naturally it moved. Participants provided a rating from 1 to 5 on each dimension using a semantic differential scale (Figure A.4).

Figure A.4. The scales above contributed to a score for anthropomorphism. The higher the rating, the more natural the participant perceived the robot to be in movement.



For this study, I selected from the entire Godspeed questionnaire with exception to items that fell under perceived safety because they were not relevant. This leaves a total score of 105 possible, given 21 separate scales. Appendix B presents the Godspeed questionnaire for only the selected dimensions used in this study.

Apparatus and Stimuli

The experiment was conducted on a commercial computer that projected to an attached 24" monitor. Experiment presentations were delivered via a PowerPoint presentation. Character models in the anthropomorphic presentations were generated using Photoshop and Crazy Talk 8 [166]. Crazy Talk 8 provided the ability to control all aspects of the face, including the blinking, eyebrow and forehead, orientation of the head, and over all affect using stored profiles (e.g., sad, happy, and worried). Mouth movements were automatically synched to synthesized voice recordings and the strength of the movements modified using a slider provided by the software's interface. Some manual puppeteering was required to remove repetitious animation that would make presentations appear less natural. Voice synthesis were generated using Amazon Web Services' Polly application. The voice to be selected was a natural British accent. A total of 252 unique animations were created for the ratings prompts and backstories.

The experiment was administered in a closed sound-proofed room under dimmed lighting (Figure A.5). The experimenter controlled the presentations from the other side of a partition that separated him from the participant. From the control area the experimenter enabled or disabled the mouse or voice recognition, as well as start and stop the presentations. Participants saw the presentations on the opposite side of the partition where the display was adjusted so that agent on the screen was at eye level.

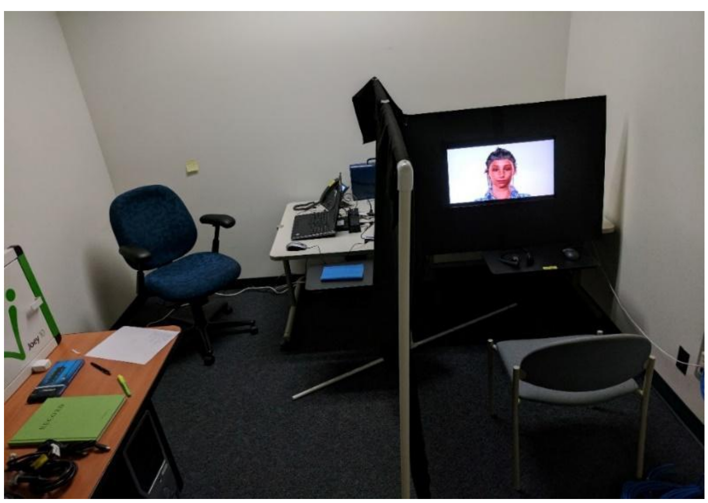


Figure A.5. In this experiment room the experimenter sat on the left of the partition and the participant viewed the presentation from a monitor on the right.

Results

Participants received twelve trials each and ten prompts for ratings within each trial. The mean number of ratings obtain from participants between the agent ($M = 7.3, SD = 3.8$) and no agent ($M = 7.3, SD = 4.0$) conditions were equivalent. Thus, no main findings for Anthropomorphism will be expected. For Interaction Method, the mouse condition ($M = 7.6, SD = 3.6$) did obtain more ratings than the voice condition ($M = 7.0, SD = 4.1$).

As a manipulation check a questionnaire was administered after each block to index the anthropomorphic effect of the Agent and No Agent condition (Table A.3). The instrument indexed anthropomorphism on two dimensions, humanness and eeriness. In addition, diminished anthropomorphism was indexed in terms of how eerie the presentation was to the participant. As expected, on the humanness dimensions, the Agent condition ($M = 13.6, SD = 8$) resulted in higher anthropomorphism scores than the No Agent condition ($M = 12.4, SD = 5.2$). Similarly, on the attractiveness dimension, the Agent condition ($M = 15.4, SD = 3.7$) resulted in higher anthropomorphism than in the No Agent condition ($M = 14.5, SD = 3.7$). However, both the Agent ($M = 24.7, SD = 7.1$) and No Agent ($M = 24, SD = 5.6$) condition resulted in almost the same level of eeriness.

Table A.2. Mean number of ratings (10 possible per trial) obtained per trial for each condition and standard deviations.

Anthropomorphism

Interaction Method

	No Agent	Agent
Mouse	M=7.6, SD=3.6	M=7.6, SD=3.7
Voice	M=7.1, SD=4.0	M=6.9, SD=4.3

Table A.3. Average anthropomorphism scores across Anthropomorphism condition

	No Agent	Agent
Humanness (Max.30)	M=12.4, SD=5.2	M=13.6, SD=8
Attractiveness (Max.40)	M=14.5, SD=3.7	M=15.4, SD=4.8
Eeriness (Max.25	M=24.0, SD=5.6	M=24.7, SD=7.1

APPENDIX B. PRE-SESSION QUESTIONNAIRE

Pre-Session Questionnaire

* Required

1. **Participant ID#** *

2. **Date** *

   Example: December 15, 2012

3. **Time** *

   Example: 8:30 AM

Interface Interaction Techniques

4. **Rate your preferences for the following interface interaction techniques.** *

   Mark only one oval per row.

	Dislike a great deal	Somewhat Dislike	Neutral	Somewhat Like	Like a great deal
Voice input	◯	◯	◯	◯	◯
Mouse input	◯	◯	◯	◯	◯
Keyboard input	◯	◯	◯	◯	◯
Animated agent	◯	◯	◯	◯	◯

5. **Rate your preferences for the following interface interaction qualities.** *

   Mark only one oval per row.

	Dislike a great deal	Somewhat Dislike	Neutral	Somewhat Like	Like a great deal
Speed	◯	◯	◯	◯	◯
Convenience	◯	◯	◯	◯	◯
Engagement	◯	◯	◯	◯	◯

Individual Differences in Anthropomorphism Questionnaire (IDAQ)

Provide a rating from 0 to 10 for the following questions, where 0 means "not at all" and 10 means "very much".

6. **To what extent is the desert lethargic?** *

   Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

7. **To what extent is the average computer average?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

8. **To what extent does technology - devices and machines for manufacturing, entertainment, and productive process (e.g., cars, computers, television sets) - have intentions?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

9. **To what extent does the average fish have free will?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

10. **To what extent are pets useful?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

11. **To what extent does the average mountain have free will?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

12. **To what extent is the average amphibian lethargic?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

13. **To what extent does a television set experience emotions?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

14. **To what extent is the average robot good-looking?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

15. **To what extent does the average robot have consciousness?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

16. **To what extent do cows have intentions?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

17. **To what extent does a car have free will?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

18. **To what extent does the ocean have consciousness?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

19. **To what extent is the average camera lethargic?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

20. **To what extent is a river useful?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

21. **To what extent does the average computer have a mind of its own?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

22. **To what extent is a tree active?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

23. **To what extent is the average kitchen appliance useful?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

24. **To what extent does a cheetah experience emotions?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

25. **To what extent does the environment experience emotions?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

26. **To what extent does the average insect have a mind of its own?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

27. **To what extent does an insect have a mind of its own?** *
Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Very much

28. **To what extent does a tree have a mind of its own?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

29. **To what extent is technology - devices and machines for manufacturing, entertainment, and productive processes (e.g.,cars, computers, television sets) - durable?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

30. **To what extent is the average cat active?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

31. **To what extent does the wind have intentions?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

32. **To what extent is the forest durable?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

33. **To what extent is a tortoise durable?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

34. **To what extent does the average reptile have consciousness?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

35. **To what extent is the average dog good-looking?** *

Mark only one oval.

	0	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Very much

The Questionnaire is now complete.

Please press submit and inform the Researcher that you have finished the questionnaire.

APPENDIX C. ANTHROPOMORPHISM INDEX

Block Questionnaire

* Required

1. **Participant ID#** *

2. **Block Number** *
   Mark only one oval.
   - 1
   - 2
   - 3
   - 4

3. **Interaction Method** *
   Mark only one oval.
   - Mouse *Skip to question 4.*
   - Voice *Skip to question 5.*

Mouse Interaction Experience

4. **Please rate your experience with the mouse interaction method.** *
   Mark only one oval.

	1	2	3	4	5	
Not preferred						Preferred

Skip to question 6.

Voice Interaction Method Experience

5. **Please rate your experience with the voice interaction method.** *
   Mark only one oval.

	1	2	3	4	5	
Not preferred						Preferred

Skip to question 6.

Humanness Index

Please rate your experience with the previous block of trials.

6. *

Mark only one oval.

	1	2	3	4	5	
Artificial	◯	◯	◯	◯	◯	Natural

7. *

Mark only one oval.

	1	2	3	4	5	
Human-made	◯	◯	◯	◯	◯	Human-like

8. *

Mark only one oval.

	1	2	3	4	5	
Without Definite Lifespan	◯	◯	◯	◯	◯	Mortal

9. *

Mark only one oval.

	1	2	3	4	5	
Inanimate	◯	◯	◯	◯	◯	Living

10. *

Mark only one oval.

	1	2	3	4	5	
Mechanical Movement	◯	◯	◯	◯	◯	Biological Movement

11. *

Mark only one oval.

	1	2	3	4	5	
Synthetic	◯	◯	◯	◯	◯	Real

Eeriness Index

Please rate your experience with the previous block of trials.

12. *

Mark only one oval.

	1	2	3	4	5	
Reassuring	◯	◯	◯	◯	◯	Eerie

13. *

Mark only one oval.

	1	2	3	4	5	
Numbing	◯	◯	◯	◯	◯	Freaky

14. *

Mark only one oval.

	1	2	3	4	5	
Ordinary	◯	◯	◯	◯	◯	Supernatural

15. *

Mark only one oval.

	1	2	3	4	5	
Bland	◯	◯	◯	◯	◯	Uncanny

16. *

Mark only one oval.

	1	2	3	4	5	
Unemotional	◯	◯	◯	◯	◯	Hair-raising

17. *

Mark only one oval.

	1	2	3	4	5	
Uninspiring	◯	◯	◯	◯	◯	Spine-tingling

18. *

Mark only one oval.

	1	2	3	4	5	
Predictable	◯	◯	◯	◯	◯	Thrilling

19. *

Mark only one oval.

	1	2	3	4	5	
Boring	◯	◯	◯	◯	◯	Shocking

Attractiveness Index

Please rate your experience with the previous block of trials.

20. *

Mark only one oval.

	1	2	3	4	5	
Unattractive	◯	◯	◯	◯	◯	Attractive

21. *

Mark only one oval.

	1	2	3	4	5	
Repulsive	◯	◯	◯	◯	◯	Agreeable

22. *

Mark only one oval.

	1	2	3	4	5	
Ugly	◯	◯	◯	◯	◯	Beautiful

23. *

Mark only one oval.

	1	2	3	4	5	
Messy	◯	◯	◯	◯	◯	Sleek

24. *

Mark only one oval.

	1	2	3	4	5	
Crude	◯	◯	◯	◯	◯	Stylish

The Questionnaire is now complete.

Please submit questionnaire and inform the research that you have finished

APPENDIX D. END OF SESSION QUESTIONNAIRE

End of Session Questionnaire

* Required

1. **Participant ID#** *

2. **Was the agent anthropomorphic or human-like?** *
   Mark only one oval.
   - Yes
   - No
   - Unsure

3. **Why or why not?** *

Comparisons

4. **Which of the input methods did you prefer?** *
   Mark only one oval.
   - Mouse
   - Voice
   - Neither
   - Unsure

5. **Which of the presentation methods did you prefer?** *
   Mark only one oval.
   - Animated voice agent
   - Synthetic voice only
   - Neither
   - Unsure

6. **Which of the presentation techniques best suited the ratings solicitation task?** *
   Mark only one oval.
   - Animated voice agent *Skip to question 7.*
   - Synthetic voice only *Skip to question 8.*
   - Neither *Skip to question 9.*
   - Unsure *Skip to question 10.*

7. **Why does the animated voice agent best suit the ratings solicitation task? Is there a technique that may work better?** *

Skip to "The Questionnaire is now complete.."

8. **Why does the synthetic voice only technique best suit the ratings solicitation task? Is there a technique that may work better?** *

Skip to "The Questionnaire is now complete.."

9. **Why doesn't either technique best suit the ratings solicitation task? Is there a technique that may work better?** *

Skip to "The Questionnaire is now complete.."

10. **Why? Is there a technique that may work better?** *

11. **If/when you chose not to provide ratings, for what reason did you decide not to do so?** *

Skip to "The Questionnaire is now complete.."

The Questionnaire is now complete.

Please submit the questionnaire and inform the researcher that you have finished.

APPENDIX E. NUMBER OF RATINGS OBTAINED FROM PARTICIPANTS ACROSS CONDITIONS

Participant ID	Voice.AVA.Mus	Voice.AVA.Bio	Voice.AVA.Sci	Keyboard.AVA.Mus
1	10	10	10	10
2	0	0	0	5
3	10	10	10	10
4	10	10	10	10
5	10	10	4	10
6	10	10	10	10
7	4	10	10	10
8	5	4	7	4
9	10	10	10	10
10	1	0	0	0
11	4	7	10	10
12	10	10	10	10
13	3	10	10	10
14	10	10	10	10
15	10	10	10	10
16	10	10	10	7
17	2	10	10	4
18	0	1	2	3
19	4	10	0	4
20	1	2	0	2
21	10	10	10	10
22	3	6	0	0
23	8	10	10	10
24	3	10	10	4
Column Totals	148	190	173	173

Participant ID	Keyboard.AVA.Bio	Keyboard.AVA.Sci	Voice.noAVA.Mus	Voice.noAVA.Bio
1	10	10	10	10
2	0	0	0	0
3	10	10	10	10
4	10	10	3	9
5	10	10	8	10
6	10	10	5	4
7	10	10	10	0

8	10	10	0	0
9	10	10	10	10
10	1	1	1	1
11	10	10	10	3
12	10	10	10	10
13	10	10	10	10
14	10	10	10	10
15	10	10	10	10
16	10	10	10	10
17	2	3	5	10
18	6	6	5	0
19	4	10	0	1
20	3	1	1	0
21	10	10	10	10
22	6	1	2	0
23	10	9	10	10
24	10	4	10	10
Column Totals	192	185	160	148

Participant ID	Voice.noAVA.Sci	Keyboard.noAVA.Mus	Keyboard.noAVA.Bio	Keyboard.noAVA.Sci
1	10	10	10	10
2	0	0	0	0
3	10	10	10	10
4	10	10	10	10
5	10	10	10	10
6	10	10	10	10
7	10	10	8	10
8	0	0	3	5
9	10	10	10	10
10	1	0	0	0
11	10	10	10	10
12	10	10	10	10
13	10	10	10	10
14	10	10	10	9
15	10	10	10	10
16	10	5	10	9
17	10	10	10	9
18	10	6	6	6
19	0	10	10	10

20	1	2	1	1
21	10	10	10	9
22	6	1	0	4
23	10	10	10	9
24	10	7	2	7
Column Totals	188	181	180	188

APPENDIX F. PRE-SESSION QUESTIONNAIRE OUTPUT

Participant ID#	or the following interface interaction t	r the following interface interaction te	for the following interface interaction techn	e following interface interaction t	s for the following interface interact	ces for the following interface interaction qual	ces for the following interface interaction qualit
25	5	4	4	5	5	5	5
26	5	4	3	5	5	5	5
27	2	4	4	3	4	3	3
28	2	3	3	2	4	4	3
29	4	4	3	3	5	5	3
30	4	5	4	3	5	4	4
31	4	3	3	4	3	4	3
32	2	4	5	2	2	3	2
33	5	3	3	4	5	4	4
34	1	5	5	1	4	4	4
35	2	4	4	2	4	2	4
36	2	4	5	3	4	5	3
37	Somewhat Like	Like a great deal	Like a great deal	Somewhat Dislike	Like a great deal	Like a great deal	Somewhat Like
38	Dislike a great deal	Like a great deal	Somewhat Like	Neutral	Somewhat Like	Neutral	Like a great deal
39	Somewhat Like	Like a great deal	Like a great deal	Somewhat Dislike	Neutral	Neutral	Neutral
40	Neutral	Somewhat Like	Somewhat Like	Somewhat Dislike	Neutral	Neutral	Somewhat Like
41	Somewhat Dislike	Like a great deal	Somewhat Like	Like a great deal	Somewhat Like	Like a great deal	Like a great deal
42	Somewhat Dislike	Neutral	Neutral	Neutral	Neutral	Neutral	Somewhat Dislike
43	Neutral	Somewhat Like	Somewhat Like	Neutral	Somewhat Like	Neutral	Somewhat Like
44	Somewhat Dislike	Somewhat Like	Like a great deal	Somewhat Like	Neutral	Somewhat Like	Like a great deal
45	Somewhat Like	Like a great deal	Like a great deal	Neutral	Like a great deal	Like a great deal	Like a great deal
46	Somewhat Dislike	Neutral	Neutral	Somewhat Dislike	Neutral	Dislike a great deal	Somewhat Dislike
47	Dislike a great deal	Like a great deal	Like a great deal	Somewhat Like	Dislike a great deal	Like a great deal	Like a great deal
48	Dislike a great deal	Somewhat Like	Neutral	Somewhat Dislike	Like a great deal	Somewhat Like	Neutral

Participant ID#	To what extent is the desert lethargic	at extent is the average computer av	ufacturing, entertainment, and productive	xtent does the average fish have	To what extent are pets useful?	at extent does the average mountain have fre	what extent is the average amphibian letharg	es a television set ex	t is the average robot	the average robot
25	8	5	7	8	9	0	6	0	1	2
26	0	6	0	10	10	10	0	0	5	0
27	7	5	8	8	7	5	4	0	0	2
28	5	6	7	3	10	0	0	0	0	5
29	5	5	7	3	7	1	4	0	1	1
30	4	7	7	5	7	1	7	0	2	3
31	4	10	5	9	10	3	1	0	5	0
32	7	2	8	2	10	3	3	0	0	0
33	7	6	8	6	8	7	4	1	5	7
34	5	5	1	0	10	0	0	0	0	0
35	10	7	9	7	10	0	4	0	2	0
36	5	10	0	3	8	0	5	0	3	0
37	5	10	10	7	8	0	5	0	0	4
38	0	10	0	0	9	0	0	0	5	0
39	5	5	8	1	9	5	5	9	6	3
40	5	5	9	5	10	10	5	0	0	4
41	5	7	9	7	6	8	9	0	8	0
42	2	5	7	3	9	5	5	8	0	0
43	5	4	7	6	8	3	5	2	4	6
44	5	7	10	8	8	7	6	6	6	7
45	0	4	0	0	8	0	3	0	0	0
46	6	6	6	4	3	5	4	4	2	4
47	5	3	7	3	8	3	5	6	10	9
48	2	5	7	0	8	0	0	0	2	0

Participant ID#	o what extent do cows have intention	o what extent does a car have free wi	what extent does the ocean have consciousn	t extent is the average camera le	To what extent is a river useful?	tent does the average computer have a mind	To what extent is a tree active?	the average kitchen	does a cheetah experi	the environment e
25	5	3	1	2	10	5	8	8	7	7
26	6	0	0	0	10	0	0	10	5	0
27	1	0	0	8	9	0	5	6	3	2
28	6	0	5	3	2	3	5	7	7	4
29	1	3	4	2	9	2	8	9	8	8
30	6	0	3	2	9	2	9	9	10	6
31	8	0	3	0	10	0	8	8	10	10
32	7	0	0	0	10	0	10	10	10	6
33	7	2	1	2	9	9	5	8	5	9
34	0	8	0	0	10	4	0	10	10	0
35	8	0	1	1	9	1	8	7	8	7
36	2	0	0	1	10	0	5	8	3	0
37	7	0	3	5	10	9	9	10	8	6
38	10	0	0	0	10	0	10	10	5	0
39	5	8	5	9	6	6	7	9	6	10
40	7	0	0	5	9	2	10	10	8	8
41	9	1	1	0	9	1	10	9	10	9
42	2	0	1	3	8	2	8	10	7	0
43	7	5	4	5	7	4	4	8	6	4
44	6	7	8	6	8	9	6	10	7	8
45	1	0	0	0	6	0	9	10	3	0
46	3	4	3	4	3	5	3	5	3	5
47	7	6	10	5	10	6	10	8	6	10
48	0	0	0	0	10	0	4	5	7	0

Participant ID#	nt does the average insect have a mi	extent does an insect have a mind of	what extent does a tree have a mind of its o	cturing, entertainment, and proc	what extent is the average cat acti	o what extent does the wind have intentions	To what extent is the forest durable?	extent is a tortoise	s the average reptile h	is the average do
25	8	8	1	7	9	0	6	6	8	10
26	3	3	0	7	4	0	8	7	1	10
27	8	8	0	7	7	0	5	5	3	5
28	6	6	3	5	6	3	3	7	7	6
29	6	4	2	7	4	6	3	8	2	5
30	5	5	4	6	7	6	7	7	9	7
31	10	10	5	8	9	0	5	7	4	5
32	10	10	2	7	10	0	4	6	10	6
33	6	8	9	7	7	5	6	6	6	8
34	10	10	0	0	10	0	10	3	0	10
35	7	8	1	8	7	0	8	7	8	10
36	3	3	0	8	7	0	6	8	10	7
37	10	9	4	6	6	5	8	8	9	10
38	0	0	0	5	5	0	5	5	0	5
39	6	6	5	8	9	7	6	6	7	7
40	6	8	10	7	6	0	4	6	7	8
41	10	9	6	7	8	6	7	8	10	7
42	4	4	6	9	10	0	1	4	4	5
43	6	6	5	5	6	5	7	7	7	7
44	7	6	7	6	6	3	10	3	5	5
45	0	0	0	2	4	0	3	2	4	0
46	4	2	3	3	4	4	4	5	4	4
47	8	8	10	8	7	5	7	7	6	8
48	2	2	0	7	3	0	6	6	0	10

APPENDIX G. ANTHROPOMORPHISM INDEX OUTPUT

Participant ID#	Total Humanness	Eeriness total	Attractiveness total	Condition
25	26	22	21	AVA
25	27	23	22	AVA
26	26	31	21	AVA
26	30	33	25	AVA
27	12	22	17	AVA
27	6	24	18	AVA
28	12	22	15	AVA
28	6	24	15	AVA
29	18	21	15	AVA
29	10	29	15	AVA
30	16	22	17	AVA
30	6	24	15	AVA
31	8	19	11	AVA
31	15	15	13	AVA
32	21	26	19	AVA
32	20	26	18	AVA
33	6	38	10	AVA
33	6	40	7	AVA
34	18	27	19	AVA
34	6	36	19	AVA
35	13	15	12	AVA
35	6	25	9	AVA
36	6	15	7	AVA
36	6	13	10	AVA
25	8	19	11	noAVA
25	8	16	11	noAVA
26	17	28	21	noAVA
26	27	24	19	noAVA
27	9	28	19	noAVA
27	13	23	15	noAVA
28	12	26	13	noAVA
28	6	19	11	noAVA
29	13	27	15	noAVA
29	20	23	16	noAVA
30	17	23	18	noAVA

30	6	19	15	noAVA
31	18	24	16	noAVA
31	12	21	15	noAVA
32	9	17	13	noAVA
32	19	31	19	noAVA
33	10	36	6	noAVA
33	10	36	7	noAVA
34	8	30	16	noAVA
34	10	17	12	noAVA
35	8	23	15	noAVA
35				noAVA
36	15	20	17	noAVA
36	11	21	14	noAVA

APPENDIX H. POST SESSION QUESTIONNAIRE OUTPUT

Participant ID#	Do you think that Ava was	Why or why not?	Which of the input method	Which of the presentation	Which of the presentation	Why does the animated vi	Why does the synthetic vc	Why doesn't either technic	Why? Is there a technique	If/when you chose not to provide ratings, for what reason did you decide not to do so?
25	Yes	Her physical features were	Voice	Animated voice agent	Animated voice agent	It feels as if you are having a conversation with someone who wants to improve the experience for you. Unlike the synthetic agent, the animated voice agent did not just ask the straight questions, but told short stories and varied her question structure to keep me engaged.				
26	Yes	Because her voice had int	Mouse	Animated voice agent	Animated voice agent	The animated voice agent is best because it's more entertaining to work with than just the voice.				
27	Yes	Because she had a lot of	Mouse	Animated voice agent	Animated voice agent	The animated voice agent is better because for one I could understand her better and two it was just interesting to look at.				
28	No	She felt uncanny and robo	Voice	Synthetic voice only	Synthetic voice only		It was faster and efficient, no			
29	Yes	there is life like feature like	Mouse	Animated voice agent	Animated voice agent	It is better to listen to than the speaker which sound too synthetic				
30	Yes	She looked like a human a	Voice	Animated voice agent	Synthetic voice only		I preferred the synthetic voice just because it was quicker and got straight to the point. I'm not sure as to what techniques might best suit the experiment better.			
31	Unsure	It felt more realistic when	Voice	Animated voice agent	Animated voice agent	I feel the animated voice agent was best for the ratings solicitation task because it made the task more interesting. It provided an opinion, but I feel that I chose to follow my own opinion regardless. When using the other methods, it became tiring more quickly than with this method.				
32	Yes	facial features and details	Voice	Animated voice agent	Animated voice agent	feels like you're talking to someone about it				
34	No	She was not fluid	Mouse	Neither	Neither			It was weird hearing an animation give instructions.		
32	Yes	Ava spoke to about perso	Mouse	Animated voice agent	Animated voice agent	It is more comfortable and initiates a sense of importance about the ratings as it feels more personal with the animated voice agent.				
35	No	Her answers sounded pre	Mouse	Synthetic voice only	Synthetic voice only		I like how it was straight to the point. The other was more creepy because it was an AI attempting to be human-like.			
36	Yes	Her conversational style is	Mouse	Unsure	Synthetic voice only		I thought the animated voice agent had some pretty engaging conversations, but it got tedious fast, especially since all I could respond with was with ratings. The synthetic voice was boring to listen to, but it allowed me to give faster/more efficient responses.			
37	Yes	She seemed very human l	Voice	Synthetic voice only	Synthetic voice only		It feels more like I am talking to a person instead of a robot			
38	Yes	She looked and responde	Mouse	Animated voice agent	Synthetic voice only		I liked responding to the synthetic voice overall only because it was straight-forward. I liked the idea of responding to Ava, however, some of her responses were predictable. I would prefer to respond to Ava but have her provide less background information and opinions, and would instead prefer for her to ask the question.			
40	Unsure	Her appearance was prett	Mouse	Animated voice agent	Synthetic voice only		I found it better because it was more straight to the point. When giving a review, I do not find it necessary to have the human interaction in between the reviews, so the synthetic voice system was more preferable for myself. I think if you used the animation and went straight to the point it would be much more enjoyable than one or the other.			
41	Yes	You could see her pores.	Mouse	Synthetic voice only	Synthetic voice only		Because they don't give you their opinion on their rating, unlike Ava			
42	No	Her movement and voice	Mouse	Synthetic voice only	Unsure				I think the voice was much	If some of the trailers were confusing or not catching my eye.
43	Yes	Ava was human-like from	Mouse	Animated voice agent	Neither			I liked Ava, I just didn't like how she talks too much. I tend to zone out when she's speaking. I liked the plain voice because it was straightforward, but I didn't like the voice since it was too robotic to the point of being almost unable to understand it.		
44	Yes	I liked it because it seeme	Mouse	Animated voice agent	Animated voice agent	The animated voice agent best suit the ratings because it was fast and because it was straight forward with all the questions that where asked about all the films. Therefore, there is no technique that may work better because it's fine.				
45	Yes	Her responses were very	Voice	Synthetic voice only	Unsure				I liked the synthetic voice	I never chose not to provide ratings.
46	Yes	she answer the responses	Mouse	Neither	Neither			mouse better, there is a bug where if I say no the slide pops up		
47	Yes	her hair flips around as sh	Voice	Animated voice agent	Animated voice agent	The voice agent gave some suggestive hints to my ratings, but it didn't affect my input				
48	Yes	She gave anecdotes abou	Mouse	Animated voice agent	Animated voice agent	She seemed to appreciate my participation more than the synthetic voice did.				

APPENDIX I. AGENT BACKSTORIES

Backstory 1

<speak>

<prosody pitch='default' range="default">

I would like you to watch this movie with me. I selected it because I thought the cinematography <prosody volume="x-soft"><emphasis level="strong">was lovely</emphasis></prosody>. After the trailer, I'd like you to tell me what you think of it. <prosody rate="medium">It will allow me to make better recommendations for you later.</prosody>**.**

</prosody>

</speak>

Backstory 2

<speak>

<prosody pitch='default' range="default">

I browsed through a bunch of other trailers</prosody> online and came across this one. It looked<break time="50ms"/> <emphasis level="moderate">pretty</emphasis>good.<break time="50ms"/> <prosody pitch="high" rate="fast">I was going to go check it out </prosody>at the theaters, <emphasis level="strong">but</emphasis> wanted get your opinion <emphasis level="strong"> too.</emphasis> Let me show it to you.</prosody>

</prosody>

</speak>

Backstory 3

<speak>

I wasn't too particularly impressed with this movie the first time I saw it, but after giving it another go I realized that there is quite a bit of depth to the story. So you might want to try watching the movie even if the trailer doesn't impress you. I'll tell you more about what I thought when I ask you to give ratings later</speak>.

Backstory 4

<speak>

I am always up for a good sci-fi. Just the other day I saw this film, Ex Machina. I'm not sure if you've seen it. <emphasis level="strong">It was a total trip.</emphasis> I won't spoil it for you, but basically it's about this robot that learns about people and figures out how to manipulate them. I think the effects alone make it worthwhile. Hang on I'll load it up and we'll watch it together.
</speak>

Backstory 5
<speak>
I hope<break time="100ms" /> I'm making this fun for you, but if you find yourself getting tired of this experiment, just remember that you don't have to give me ratings if you don't want to, and you can just skip through. If instead <break time="100ms" /> you like watching these trailers with me, then by all means keep giving me ratings so that I can find better films for you. Now this next one is one of my favorites. What I like about it most is the setting, but I won't spoil it for you. Let's check it out!
</speak>

Backstory 6
<speak>
Most people find biographies too dry, but I rather enjoy them</prosody><break time="100ms" />because the lives of the people in those films made great stories in themselves. Nowadays, they make some really good biographies with all the new flashy cinema tech. Let me show this one to you, so I can get your ratings biographies. I hope you enjoy it as much as I did.
</speak>

APPENDIX J. AGENT REQUEST SCRIPT

Number	Dimension	Set	Script
1	Acting	1	I wasn't fond about the acting, but how many stars would you give it?
2	Acting	2	The acting seemed a bit mediocre to me, but the main actor seemed to make up for the rest of the cast. What do you think about this one?
3	Acting	3	What you think about the acting? I liked it.
4	Acting	4	Can you tell me what you thought about the acting?
5	Acting	5	How many stars you want to give the acting?
6	Acting	6	I really don't know what to make of the acting. Maybe you can tell me what you think.
7	Casting	1	I saw a lot of familiar faces, but how about you? Did you like the cast?
8	Casting	2	I am not the best judge of acting, but it seemed underwhelming to me in this film. What is your rating for this one?
9	Casting	3	And the casting?
10	Casting	4	How about the casting?
11	Casting	5	Nice, what about the casting?
12	Casting	6	Tell me what you think about the casting.
13	Cinema	1	Ok, so tell me what you think about the cinematography.
14	Cinema	2	Alright, now how do you feel about the cinematography?
15	Cinema	3	Ok, now give me a rating for the cinematography.
16	Cinema	4	I think they did some great camera work on this one. What did you think about the cinematography?
17	Cinema	5	So, how did you like the cinematography?
18	Cinema	6	Great. Now, what did you think about the cinematography?
19	Costumes	1	I wasn't feeling the costumes, but what about you?
20	Costumes	2	The costumes seem out of place for me. What did you think about the costumes?
21	Costumes	3	And the costumes?
22	Costumes	4	How about the costumes?
23	Costumes	5	How many stars do you want to give the costumes?
24	Costumes	6	Costumes looked good to me in this one. How many stars do you want to give it?
25	Plot	1	Very good. I am starting to get a good feel for your taste in films. Now, how would you rate the plot?
26	Plot	2	The plot seemed underwhelming for me. How about you?
27	Plot	3	I couldn't tell from the trailer where they were going with the plot. What about you?
28	Plot	4	Plots are the most important thing to consider in considering a movie, I think. What would you rate the plot here?
29	Plot	5	And what about the plot?
30	Plot	6	Nice, now what did you think about the plot?
31	Screen	1	Tell me what you think about the screen play.
32	Screen	2	And what about the screen play?
33	Screen	3	How about that screen play?
34	Screen	4	I really didn't know what to think about that screen play. Perhaps you have an opinion? What's your rating for it?
35	Screen	5	I never really had an appreciation for screen play growing up. What would you rate it though?
36	Screen	6	Now, what's your rating for screen play?
37	Setting	1	What about the setting? How many ratings would you give that?
38	Setting	2	I have a strong preference for period dramas. What would you give for the setting in this film?
39	Setting	3	How about the setting?
40	Setting	4	Excellent, let's get a rating for the setting then.
41	Setting	5	Great, give me a rating for the setting, please.
42	Setting	6	And the setting?
43	Soundtrack	1	How about the soundtrack?
44	Soundtrack	2	I love a good soundtrack. What is your rating for the soundtrack in this film?
45	Soundtrack	3	What is your rating for the soundtrack?
46	Soundtrack	4	How would you rate the soundtrack?
47	Soundtrack	5	And the soundtrack?
48	Soundtrack	6	What about the soundtrack? What is your rating for that?
49	Story line	1	And how about the story line?
50	Story line	2	What about that storyline? I thought the build up was a bit slow, but what about you?
51	Story line	3	I loved the story line in this one, but don't let me sway you. What is your honest opinion?
52	Story line	4	Wasn't too fond of the storyline, but that's me. What about you?
53	Story line	5	Great, give me a rating for the story line, please.
54	Story line	6	And the story line?
55	Style	1	And the style?
56	Style	2	I am far more concerned about the plots, but I want to know what you think of the style.
57	Style	3	Tell me what your rating is for how they styled the film.
58	Style	4	Styling is kind of vague to me, but I think it is referring to the overall aesthetic of the film. How many stars here?
59	Style	5	What about the style?
60	Style	6	Good, I am learning a lot about your tastes. How would you rate the style?

REFERENCES

[1] J. Herlocker, J. Konstan, and J. Riedl, “Explaining collaborative filtering recommendations,” *Proceedigns ACM 2000 Conf. Comput. Support. Coop. Work*, pp. 241–250, 2000.

[2] J. Ben Schafer, D. Frankowski, J. Herlocker, and S. Sen, “Collaborative Filtering Recommender Systems,” in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin/Heidelberg: Springer-Verlag, 2007, pp. 291–324.

[3] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, “Methods and Metrics for Cold-Start Recommendations,” in *SIGIR ’02 Proceedings of the 25th annual International ACM SIGIR Conference on Research and Development in Information Retreival*, 2002, pp. 253–260.

[4] D. Poirier, F. Fessant, and I. Tellier, “Reducing the cold-start problem in content recommendation through opinion classification,” *Proc. - 2010 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2010*, vol. 1, pp. 204–207, 2010.

[5] G. Guo, J. Zhang, and D. Thalmann, “Merging trust in collaborative filtering to alleviate data sparsity and cold start,” *Knowledge-Based Syst.*, vol. 57, pp. 57–68, 2014.

[6] L. H. Son, “Dealing with the new user cold-start problem in recommender systems: A comparative review,” *Inf. Syst.*, vol. In Press, pp. 1–18, 2014.

[7] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, “Facing the cold start problem in recommender systems,” *Expert Syst. Appl.*, vol. 41, no. 4 PART 2, pp. 2065–2073, 2014.

[8] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.

[9] P. Pu and L. Chen, “Trust-inspiring explanation interfaces for recommender systems,” *Knowledge-Based Syst.*, vol. 20, no. 6, pp. 542–556, Aug. 2007.

[10] T. Higgins and W. Boston, “The battle for the last unconquered screen - the one in your car,” *The Wall Street Journal*, 2019. [Online]. Available: https://www.wsj.com/articles/the-battle-for-the-last-unconquered-screenthe-one-in-your-car-11554523220. [Accessed: 04-Jun-2020].

[11] C. Nass, “Machines and Mindlessness: Social Responses to Computers.,” *J. Soc. Issues*, vol. 56, no. 1, pp. 81–104, 2000.

[12] T. Bickmore, L. Pfeifer, and D. Schulman, “Relational Agents Improve Engagement and Learning in Science Museum Visitors,” in *Intelligent Virtual Agents*, 2011, pp. 55–67.

[13] O. Perski, A. Blandford, R. West, and S. Michie, “Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis,” *Transl. Behav. Med.*, vol. 7, no. 2, pp. 254–267, 2017.

[14] H. L. O’Brien and E. G. Toms, “What is user engagement? A conceptual framework for defining user engagement with technology.,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 938–955, 2008.

[15] T. Bickmore W., S. Consolvo, and S. Intille S., “Engagement by design,” *CHI EA ’09*. ACM, pp. 4807–4810, 2009.

[16] P. Pu, L. Chen, and R. Hu, “Evaluating recommender systems from the user’s perspective: survey of the state of the art,” *User Model. User-adapt. Interact.*, vol. 22, no. 4–5, pp. 317–355, 2012.

[17] J. L. Harman, J. O’Donovan, T. Abdelzaher, and C. Gonzalez, “Dynamics of human trust in recommender systems Dynamics of Human Trust in Recommender Systems,” in *Proceedings of the 8th Conference on Recommender Systems*, 2014, no. JANUARY.

[18] P. Victor, D. C. Martine, and C. Cornelis, “Trust and Recommendations,” in *Recommender Systems Handbook2*, 1st ed., F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. New York: Springer-Verlag, 2010, pp. 645–675.

[19] F. Ricci, L. Rokach, and B. Shapira, “Introduction to Recommender Systems Handbook,” in *Recommender Systems Handbook*, New York Dorderecht Heidelberg London: Springer, 2011, pp. 1–35.

[20] J. Leino, “Juha Leino User Factors in Recommender Systems : Case Studies in e-Commerce , News Recommending , and e-Learning,” 2014.

[21] B. Smyth, “Case-based recommendation,” in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Berlin/Heidelberg: Springer-Verlag, 2007, pp. 342–376.

[22] M. Balabanović and Y. Shoham, “Fab: content-based, collaborative recommendation,” *Commun. ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[23] J. Herlocker, J. Konstan, and J. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–23, 2004.

[24] N. Tintarev and J. Masthoff, “Designing and evaluating explanations for recommender systems,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. New York: Springer, 2011, pp. 479–507.

[25] J. Tang, X. Hu, and H. Liu, *Social recommendation: a review*, vol. 3, no. 4. 2013.

[26] Y. Koren, “The bellkor solution to the netflix grand prize,” *Netflix Prize Doc.*, vol. 81, 2009.

[27] A. Toscher, M. Jahrer, and R. M. Bell, “The bigchaos solution to the netflix grand prize,” *Netflix Prize Doc.*, pp. 1–52, 2009.

[28] M. Piotte and M. Chabbert, “The pragmatic theory solution to the netflix grand prize,” *Netflix Prize Doc.*, 2009.

[29] G. Adomavicius, B. Mobasher, F. Ricci, and A. Tuzhilin, “Context-Aware Recommender Systems,” *AI Magazine*, Jun-2011.

[30] A.-Q. V Dao *et al.*, “Evaluation of a Recommender System for Single Pilot Operations,” *Procedia Manuf.*, vol. 3, pp. 3070–3077, 2015.

[31] E. Bouzekri *et al.*, “Engineering issues related to the development of a recommender system in a critical context: Application to interactive cockpits,” *Int. J. Hum. Comput. Stud.*, vol. 121, no. May 2018, pp. 122–141, 2019.

[32] L. H. Son, “HU-FCF++: A novel hybrid method for the new user cold-start problem in recommender systems,” *Eng. Appl. Artif. Intell.*, vol. 41, pp. 207–222, 2015.

[33] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User Model. User-adapt. Interact.*, vol. 12, no. 4, pp. 331–370, 2002.

[34] D. W. Oard and J. Kim, “Implicit feedback for recommender systems,” in *Proceedings of the AAAI Workshop on Recommender Systems*, 1998, pp. 81–83.

[35] L. Safoury and A. Salah, “Exploiting User Demographic Attributes for Solving Cold-Start Problem in Recommender System,” *Lect. Notes Softw. Eng.*, vol. 1, no. 3, pp. 303–307, 2013.

[36] P. Dourish, “What we talk about when we talk about context,” *Pers. Ubiquitous Comput.*, vol. 8, no. 1, pp. 19–30, 2004.

[37] M. IJtsma, L. M. Ma, A. R. Pritchett, and K. M. Feigh, “Computational Methodology for the Allocation of Work and Interaction in Human-Robot Teams,” *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 4, pp. 221–241, 2019.

[38] J. Lee, V. Venktatraman, J. L. Campbell, and C. M. Richard, “Workload and Attention Management in Automated Vehicles,” in *Human Performance in Automated and Autonomous Systems*, M. Mouloua and P. A. Hancock, Eds. Boca Raton, FL: Taylor & Francis Group, 2020, pp. 213–230.

[39] Y. Moon, “Intimate Exchanges: Using Computers to Elicit SelfDisclosure From Consumers,” *J. Consum. Res.*, vol. 26, no. 4, pp. 323–339, 2000.

[40] J. P. Bliss, G. Schwark, S. Rice, and K. Geels, “Automation Trust and Situational Experience: Theoretical Assumptions and Experimental Evidence,” in *Human Performance in Automated and Autonomous Systems*, M. Mouloua and P. A. Hancock, Eds. Boca Raton, FL, 2020, pp. 155–173.

[41] G. Sadler *et al.*, “Effects of Transparency on Pilot Trust and Agreement in the Autonomous Constrained Flight Planner,” pp. 1–9, 2016.

[42] J. B. Lyons, “Being Transparent about Transparency : A Model for Human-Robot Interaction,” in *Trust and Autonomous Systems: Papers from the 2013 AAAI Spring Symposium*, 2013, pp. 48–53.

[43] H. H. Clark, *Using language*. 1996.

[44] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. Van Riemsdijk, and M. Sierhuis, “Coactive Design: Designing Support for Interdependence in Joint Activity,” *J. Human-Robot Interact.*, vol. 3, no. 1, p. 43, 2014.

[45] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, “Ten challenges for making automation a ‘team player’ in joint human-agent activity,” *IEEE Intell. Syst.*, vol. 19, no. 6, pp. 91–95, 2004.

[46] J. Gao, J. D. Lee, and Y. Zhang, “A dynamic model of interaction between reliance on automation and cooperation in multi-operator multi-automation situations,” *Int. J. Ind. Ergon.*, vol. 36, no. 5, pp. 511–526, 2006.

[47] T. C. Skelly, K. Fries, B. Linnett, C. Nass, and B. Reeves, “Seductive interfaces: Satisfying a mass audience,” *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 1994-April, pp. 359–360, 1994.

[48] W. Quesenbury, "The five dimensions of usability," in *Content and Complexity: Information Design in Technical Communications*, M. J. Albers and M. M. B., Eds. 2002, pp. 81–102.

[49] Global Cycling Network, *Inside POC | Cycling Helmet And Clothing Design At POC HQ*. Global Cycling Network, 2018.

[50] S. H. Fairclough, K. Gilleade, K. C. Ewing, and J. Roberts, "Capturing user engagement via psychophysiology: Measures and mechanisms for biocybernetic adaptation," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 6, no. 1, pp. 63–79, 2013.

[51] M. C. Rozendaal, D. V. Keyson, and H. de Ridder, "Product features and task effects on experienced richness, control and engagement in voicemail browsing," *Pers. Ubiquitous Comput.*, vol. 13, no. 5, pp. 343–354, 2009.

[52] N. Tintarev, "Explaining recommendations," University of Aberdeen, 2009.

[53] R. Sinha and K. Swearingen, "The Role of Transparency in Recommender Systems," in *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 2000, pp. 830–831.

[54] S. Dow, M. Mehta, E. Harmon, B. MacIntyre, and M. Mateas, "Presence and engagement in an interactive drama," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 1475–1484, 2007.

[55] H. Lejeune and J. H. Wearden, "Vierordt's The Experimental Study of the Time Sense (1868) and its legacy," *Eur. J. Cogn. Psychol.*, vol. 21, no. 6, pp. 941–960, Sep. 2009.

[56] Y. Douglas and A. Hargadon, "The Pleasure Principle : Immersion , Engagement , Flow," pp. 153–160, 2000.

[57] M. I. Hwang and R. G. Thorn, "The effect of user engagement on system success: A meta-analytical integration of research findings," *Inf. Manag.*, vol. 35, no. 4, pp. 229–236, 1999.

[58] C. E. Short, A. L. Rebar, R. C. Plotnikoff, and C. Vandelanotte, "Designing engaging online behaviour change interventions: A proposed model of user engagement," *Eur. Heal. Psychol.*, vol. 17, no. 1, pp. 32–38, 2013.

[59] J. D. Lee, "Perspectives on Automotive Automation and Autonomy," *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 53–57, 2018.

[60] R. West and S. Michie, *A guide to development and evaluation of digital interventions in healthcare*. London: Silverback Publishing.

[61] D. Szafir and B. Mutlu, “Pay attention! Designing adaptive agents that monitor and improve user engagement,” *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 11–20, 2012.

[62] K. Doherty and G. Doherty, “Engagement in HCI: Conception, theory and measurement,” *ACM Comput. Surv.*, vol. 51, no. 5, 2019.

[63] M. Seah and P. Cairns, “From Immersion to Addiction in Videogames,” *Proc. 22nd Br. HCI Gr. Annu. Conf. People Comput. Cult. Creat. Interact. (BCS-HCI’08),* vol. 1, pp. 55–63, 2008.

[64] J. Freyne, M. Jacovi, I. Guy, and W. Geyer, “Increasing engagement through early recommender intervention,” *RecSys’09 - Proc. 3rd ACM Conf. Recomm. Syst.*, pp. 85–92, 2009.

[65] M. J. Kelley and N. Clausen-Grace, “Facilitating Engagement by Differentiating Independent Reading,” *Read. Teach.*, vol. 63, no. 4, pp. 313–318, 2009.

[66] Y. H. Kim, D. J. Kim, and K. Wachter, “A study of mobile user engagement (MoEN): Engagement motivations, perceived value, satisfaction, and continued engagement intention,” *Decis. Support Syst.*, vol. 56, no. 1, pp. 361–370, 2013.

[67] B. Mallon and B. Webb, “Structure, causality, visibility and interaction: Propositions for evaluating engagement in narrative multimedia,” *Int. J. Hum. Comput. Stud.*, vol. 53, no. 2, pp. 269–287, 2000.

[68] F. Biocca, C. Harms, and J. K. Burgoon, “Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria,” *Presence Teleoperators Virtual Environ.*, vol. 12, no. 5, pp. 456–480, 2003.

[69] E. Goffman, *The presentation of self in everyday life*. Garden City, NY, 1959.

[70] C. Nass and K. M. Lee, “Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction.,” *J. Exp. Psychol. Appl.*, vol. 7, no. 3, pp. 171–181, 2001.

[71] E. J. Lee, “The more humanlike, the better? How speech type and users’ cognitive style affect social responses to computers,” *Comput. Human Behav.*, vol. 26, no. 4, pp. 665–672, 2010.

[72] N. Pfeuffer, A. Benlian, H. Gimpel, and O. Hinz, “Anthropomorphic Information Systems,” *Bus. Inf. Syst. Eng.*, vol. 61, no. 4, pp. 523–533, 2019.

[73] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore, "BEAT: the Behavior Expression Animation Toolkit," *Proc. 28th Annu. Conf. Comput. Graph. Interact. Tech.*, vol. 137, no. August, pp. 477–486, 2001.

[74] T. Bickmore, D. Schulman, and L. Yin, "Maintaining engagement in long-term interventions with relational agents," *Appl. Artif. Intell.*, vol. 24, no. 6, pp. 648–666, 2010.

[75] L. Ring, D. Utami, and T. Bickmore, "The Right Agent for the Job?," in *Intelligent Virtual Agents SE - 49*, vol. 8637, T. Bickmore, S. Marsella, and C. Sidner, Eds. Springer International Publishing, 2014, pp. 374–384.

[76] E. Moriuchi, "An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior," *Psychol. Mark.*, no. August, pp. 1–22, 2020.

[77] K.-H. Yoo and U. Gretzel, "Creating More Credible and Persuasive Recommender Systems: The Influence of Source Characteristics on Recommender Systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. New York/Dorderecht/Heidelberg/London: Springer, 2011, pp. 455–477.

[78] L. Qiu, "Designing social interactions with animated avatars and speech output for product recommendation agents in electronic commerce," *ProQuest Diss. Theses*, vol. NR19903, no. March, pp. 176-176 p., 2006.

[79] L. Qiu and I. Benbasat, "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *J. Manag. Inf. Syst.*, vol. 25, no. 4, pp. 145–182, 2009.

[80] K.-H. Yoo and U. Gretzel, "The influence of virtual representatives on recommender system evaluation," in *Proceedings of the Fifteenth Americas Conference on Information Systems*, 2009, pp. 1–10.

[81] A. L. Baylor and J. Ryu, "The effects of image and animation in enhancing pedagogical agent persona," *J. Educ. Comput. Res.*, vol. 28, no. 4, pp. 373–394, 2003.

[82] P. Murano and P. O. Holt, "Evaluation of human-like anthropomorphism in the context of online bidding and affordances," *J. Comput.*, vol. 3, no. 6, pp. 2–8, 2011.

[83] J. Forlizzi, J. Zimmerman, V. Mancuso, and S. Kwak, "How Interface Agents Affect Interaction Between Humans and Computers," *DPPI '07 Proc. 2007 Conf. Des. pleasurable Prod. interfaces*, no. August, pp. 209–221, 2007.

[84] R. Pak, N. Fink, M. Price, B. Bass, and L. Sturre, “Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults.,” *Ergonomics*, vol. 55, no. 9, pp. 1059–72, Jan. 2012.

[85] T. Bickmore and J. Cassell, “Relational agents: a model and implementation of building user trust,” *CHI ’01*. ACM, pp. 396–403, 2001.

[86] L. Sproull, M. Subramani, S. Kiesler, J. H. Walker, and K. Waters, “When the Interface Is a Face,” *Human–Computer Interact.*, vol. 11, no. 2, pp. 97–124, 1996.

[87] J. N. Bailenson, D. Merget, and R. Schroeder, “The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction.,” *Presence Teleoperators Virtual Environ.*, vol. 15, no. 4, pp. 359–373, 2006.

[88] M. Moundridou and M. Virvou, “Evaluation of persona effect of an interface agent in a tutoring system,” *Jounral Comput. Assited Learn.*, vol. 18, no. 3, pp. 253–261, 2002.

[89] Y. D. Wang and H. H. Emurian, “An overview of online trust: concepts, elements and implications,” *Comput. Human Behav.*, vol. 21, no. 1, pp. 105–125, 2005.

[90] W. Wang and I. Benbasat, “Trust in and Adoption of Online Recommendation Agents,” *J. Assoc. Inf. Syst.*, vol. 6, no. 3, pp. 72–101, 2005.

[91] J. H. Walker, L. Sproull, and R. Subramani, “Using a Human Face in an Interface.,” in *Proceedings of the Conference on Human Factors in Computers*, 1994, pp. 85–91.

[92] D. M. Dehn and S. van Mulken, “The Impact of Animated Interface Agents: A Review of Empirical Research,” *Int. J. Hum. Comput. Stud.*, vol. 52, pp. 1–22, 2000.

[93] A. M. Turing, “Computing Machinery and Intelligence,” *Mind*, vol. 59, no. 236, pp. 433-460 CR-Copyright © 1950 Oxford Univers, Oct. 1950.

[94] D. Proudfoot, “Anthropomorphism and AI: Turings much misunderstood imitation game,” *Artif. Intell.*, vol. 175, no. 5–6, pp. 950–957, 2011.

[95] R. B. Cialdini, *Influence : science and practice*, 5th ed.. Boston: Boston : Pearson/Allyn & Bacon, 2009.

[96] C. T. Hill and D. E. Stull, “Disclosure Reciprocity: Conceptual and Measurement Issues,” *Soc. Psychol. Q.*, vol. 45, no. 4, pp. 238–244, 1982.

[97] D. R. Shaffer and M. M. Tomarelli, “When Public and Private Self-Foci Clash: Self-Consciousness and Self-Disclosure Reciprocity During the Acquaintance Process,” *J. Pers. Soc. Psychol.*, vol. 56, no. 5, pp. 765–776, 1989.

[98] I. Altman, *Social penetration: the development of interpersonal relationships*. New York: New York, Holt, Rinehart and Winston, 1973.

[99] M. L. Caltabiano and M. Smithson, “Variables Affecting the Perception of Self-Disclosure Appropriateness,” *J. Soc. Psychol.*, vol. 120, no. 1, pp. 119–128, 1983.

[100] V. J. Derlega, Ed., *Self-disclosure*. Newbury Park: Newbury Park : Sage Publications, 1993.

[101] J. C. Zimmer, R. Arsal, M. Al-Marzouq, D. Moore, and V. Grover, “Knowing your customers: Using a reciprocal relationship to enhance voluntary information disclosure,” *Decis. Support Syst.*, vol. 48, no. 2, pp. 395–406, 2010.

[102] P. Brown and S. C. Levinson, *Universals in language use: politeness phenomena*. Cambridge University Press, 1978.

[103] G. Kohs, *AlphaGo*. 2017.

[104] J. Cassell and T. Bickmore, “Negotiated collusion: Modeling social language and its relationship effects in intelligent agents,” *User Model. User-Adapted Interact.*, vol. 13, no. 1–2, pp. 89–132, 2003.

[105] T. Koda and P. Maes, “Agents with faces: the effect of personification,” *Proc. 5th IEEE Int. Work. Robot Hum. Commun. RO-MAN’96 TSUKUBA*, pp. 189–194, 1996.

[106] S. Kang and J. Gratch, “The effect of avatar realism of virtual humans on self-disclosure in anonymous social interactions,” *CHI EA ’10*. ACM, pp. 3781–3786, 2010.

[107] J. D. Lee and K. A. See, “Trust in Automation : Designing for Appropriate Reliance,” *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[108] J. E. Mercado, M. A. Rupp, J. Y. C. Chen, M. J. Barnes, D. Barber, and K. Procci, “Intelligent Agent Transparency in Human-Agent Teaming for Multi-UxV Management,” *Hum. Factors*, vol. 58, no. 3, pp. 401–415, 2016.

[109] Y. Seong and A. M. Bisantz, “The impact of cognitive feedback on judgment performance and trust with decision aids,” *Int. J. Ind. Ergon.*, vol. 38, no. 7, pp. 608–625, 2008.

[110] J. Y. C. Chen, M. Boyce, J. Wright, K. Procci, and M. Barnes, “SA-based Agent Transparency,” *ARL Tech. Rep.*

[111] S. F. Scallen and P. A. Hancock, “Implementing Adaptive Function Allocation,” *Int. J. Aviat. Psychol.*, vol. 11, no. 2, pp. 197–221, 2001.

[112] T. B. Sheridan, “Function allocation: algorithm, alchemy or apostasy?,” *Int. J. Hum. Comput. Stud.*, vol. 52, no. 2, pp. 203–216, 2000.

[113] N. Jordan, “Allocation of functions between man and machines in automated systems.,” *J. Appl. Psychol.*, vol. 47, no. 3, pp. 161–165, 1963.

[114] J. Annett, “Hierarchical Task Analysis,” pp. 17–35, 2003.

[115] R. Parasuraman and V. Riley, “Humans and automation: use, misuse, disuse, abuse,” *Hum. Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[116] P. M. Fitts, “Human engineering for an effective air navigation and traffic control system,” Columbus, OH, 1951.

[117] E. M. Roth, C. Sushereba, L. G. Militello, J. Diiulio, and K. Ernst, “Function Allocation Considerations in the Era of Human Autonomy Teaming,” *J. Cogn. Eng. Decis. Mak.*, vol. 13, no. 4, pp. 199–220, 2019.

[118] T. B. Sheridan and W. L. Verplank, “Human and Computer Control of Undersea Teleoperators,” Washington, DC, 2015.

[119] T. B. Sheridan, “Comments on ‘Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation’ by David B. Kaber,” *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 25–28, 2018.

[120] J. Y. C. Chen and M. J. Barnes, “Human - Agent teaming for multirobot control: A review of human factors issues,” *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 1, pp. 13–29, 2014.

[121] D. B. Kaber, “Reflections on Commentaries on ‘Issues in Human–Automation Interaction Modeling,’” *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 86–93, 2018.

[122] M. Luck, M. D’Inverno, and S. Munroe, “Autonomy: Variable and Generative,” pp. 11–28, 2003.

[123] J. M. Bradshaw, R. R. Hoffman, D. D. Woods, and M. Johnson, “The seven deadly myths of ‘autonomous systems,’” *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 54–61, 2013.

[124] T. Fong, “Autonomous Systems: NASA Capability Overview,” 2018. [Online]. Available: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20180007804.pdf.

[125] E. Salas, T. L. Dickinson, S. Converse, and S. I. Tannenbaum, “Toward an understanding of team performance and training,” in *Teams: Their training and performance*, R. W. Swezey and E. Salas, Eds. Norwood, NJ: Ablex, 1992, pp. 3–29.

[126] G. Klein, “Features of team coordination,” in *New trends in cooperative activities*, M. McNeese, M. R. Endsley, and E. Salas, Eds. Santa Monica, CA: HFES, 2001, pp. 68–95.

[127] T. Fong, “Collaborative Control : A Robot-Centric Model for Vehicle Teleoperation,” *Aaai*, no. Mishkin, p. 198, 2001.

[128] J. M. Bradshaw *et al.*, “Toward trustworthy adjustable autonomy and mixed-initiative interaction in KAoS,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3577 LNAI, no. June 2014, pp. 18–42, 2005.

[129] M. Pacaux-Lemoine and F. Vanderhaegen, “Towards Levels of Cooperation,” in *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013, pp. 291–296.

[130] P. Millot and M. P. Pacaux-Lemoine, *A common work space for a mutual enrichment of human-machine cooperation and team-situation awareness*, vol. 12, no. PART 1. IFAC, 2013.

[131] M. R. Endsley, “From Here to Autonomy: Lessons Learned from Human-Automation Research,” *Hum. Factors*, vol. 59, no. 1, pp. 5–27, 2017.

[132] D. B. Kaber, “Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation,” *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 7–24, 2018.

[133] C. D. Wickens, “Effort in Human Factors Performance and Decision Making,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 56, pp. 1329–1336, 2014.

[134] G.-H. Gil, D. Kaber, K. Kaufmann, and S.-H. Kim, “Effects of Modes of Cockpit Automation on Pilot Performance and Workload in a Next Generation Flight Concept of Operation,” *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 22, no. 5, pp. 395–406, 2012.

[135] W. W. Wierwille and S. A. Connor, “Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 25, no. 1, pp. 1–16, 1983.

[136] B. H. Kantowitz, S. G. Hart, and M. R. Bortolussi, “Measuring pilot workload in a moving-base simulator. I Asynchronous secondary choice-reaction task,” *Hum. Factors Soc. Annu. Meet. 27th*, vol. 1, pp. 319–322, 1983.

[137] S. G. Hart, “Measuring pilot workload in a motion base simulator: III synchronous secondary task,” *Hum. Factors Ergon. Soc. Annu. Meet. Proc.*, vol. 31, no. 7, pp. 834–837, 1987.

[138] M. R. Endsley and E. O. Kiris, “The out-of-the-loop performance problem and level of control in automation,” *Hum. Factors*, vol. 37, no. 2, pp. 381–394, 1995.

[139] R. J. Shively, J. Lachter, S. L. Brandt, M. Matessa, V. Battiste, and W. W. Johnson, “Why Human-Autonomy Teaming?,” in *Advances in Neuroergonomics and Cognitive Engineering*, 2018, pp. 3–11.

[140] M. R. Endsley, “Human-Automation Interaction and the Challenge of Maintaining Situation Awareness in Future Autonomous Vehicles,” in *Human Performance in Automated and Autonomous Systems*, Mouloua, Mustapha and Hancock, Peter A., Eds. Boca Raton, FL: Taylor & Francis Group, 2020, pp. 151–167.

[141] G. Tokadli and M. C. Dorneich, “Interaction Paradigms : from Human-Human Teaming to Human-Autonomy Teaming,” 2019.

[142] C. L. Baldwin and I. McCandliss, “Attention Management in Highly Autonomous Driving,” in *Human Performance in Automated and Autonomous Systems: Current Theory and Methods*, M. Mouloua and P. A. Hancock, Eds. Boca Raton, FL: Taylor & Francis Group, 2020, pp. 231–247.

[143] T. Victor *et al.*, “Analysis of naturalistic driving study data: safe glances, driver inattention, and crash risk,” Washington, DC, 2015.

[144] M. Schwarz, *Look Who’s Driving*. USA: PBS, 2019.

[145] “Collision between vehicle controlled by developmental automated driving system and pedestrian Tempe, Arizona,” Washington, DC, 2019.

[146] R. Baldwin, “NTSB Rules Autopilot was probably cause of fatal 2018 tesla model x crash,” *Car and Driver*, 2020.

[147] N. E. Boudette, “Autopilot cited in death of chinese tesla driver,” *The New York Times*, New York, 14-Sep-2016.

[148] F. Trapsilawati, C. Wickens, C. Chen, and X. Qu, “Transparency and Conflict Resolutation Automation Reliability in Air Traffic Control,” *Proc. Int. Symp. Aviat. Psychol.*, pp. 419–424, 2016.

[149] C. Wickens, “Automation Stages & Levels, 20 Years After,” *J. Cogn. Eng. Decis. Mak.*, vol. 12, no. 1, pp. 35–41, 2018.

[150] A. J. Van Rooij, T. M. Schoenmakers, A. A. Vermulst, R. J. J. M. Van Den Eijnden, and D. Van De Mheen, “Online video game addiction: identification of addicted adolescent gamers,” *Addiction*, vol. 106, no. 1, pp. 205–212, Jan. 2011.

[151] P. Wallis, “Believable conversational agents: introducing the intention map,” in *Creating Bonds with Humanoids. In Proc. of the Workshop at AAMAS*, 2005.

[152] L. N. Foner, “Are we having fun yet? Using social agents in social domains,” in *Human Cognition and Social Agent Technology*, K. Dauthenhahn, Ed. John Benjamins, 2000.

[153] A. De Angeli, G. I. Johnson, and L. Coventry, “The unfriendly user: exploring social reactions to chatterbots,” *Proc. Int. Conf. Affect. Hum. Factors Des. London*, pp. 467–474, 2001.

[154] P. Maes, “How to do the right thing,” Cambridge, MA, 1989.

[155] E. Goffman, “On Face-Work,” *Psychiatry (New York)*, vol. 18, no. 3. pp. 213–231, 1955.

[156] X. Amatriain, J. Pujol, N. Tintarev, and N. Oliver, “Rate it again: increasing recommendation accuracy by user re-rating,” *RecSys ’09 Proc. third ACM Conf. Recomm. Syst.*, pp. 173–180, 2009.

[157] P. O’Connor and R. Flin, “Crew Resource Management training for offshore oil production teams,” *Saf. Sci.*, vol. 41, no. 7, pp. 591–609, 2003.

[158] M. A. Ricci and J. R. Brumsted, “Crew resource management: Using aviation techniques to improve operating room safety,” *Aviat. Sp. Environ. Med.*, vol. 83, no. 4, pp. 441–444, 2012.

[159] J. K. Lauber, “Resource Management on the Flight Deck,” *Airl. Pilot*, vol. 53, pp. 20–23, 1984.

[160] B. G. Kanki, R. L. Helmreich, and J. Anca, *Crew Resource Management*, 2nd ed. San Diego, CA: Academic Press, 2010.

[161] J. M. Orasanu, “Shared mental models and crew decision making,” Princeton, NJ, 1990.

[162] H. C. Foushee, J. K. Lauber, M. M. Baetge, and D. B. Acomb, “Crew factors in flight operations III: The operational significance of exposure to short-haul air transport operations,” *NASA Tech. Memo.*, vol. 88322, no. August 1986, p. 63, 1986.

[163] M. Mori, “Bukimi no tani (the uncanny valley),” *Energy*, vol. 7, no. 4, pp. 33–5, 1970.

[164] M. Siegel, C. Breazeal, and M. I. Norton, “Persuasive Robotics: The influence of robot gender on human behavior,” *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*. pp. 2563–2568, 2009.

[165] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, 2009.

[166] “Crazy Talk 8.” .