Ambiguity Aversion in HRI Collaborative Tasks

Jordan Jackson
Department of Computer and Information Science
Cornell University
jcj65@cornell.edu

Abstract

Understanding the transfer of human decision theory models to human-robotic collaborative interactions has been a critical research questions driving the design and scalability of HRI systems. This is especially crucial as robots transition from utility tools to partners in explorative research and teamwork. In this work, I explore a crowdsourced-data approach to understanding the Ellsberg *risk in ambiguity aversion* model and how it applies to human-robot collaborative teamwork on Amazon's Mechanical Turk platform. I'll present a replicated human-robotic parallel collaborative game in which participants were asked to decide which of their two teammates — one human, one robot — were to go to a set of rooms, differing on conditions of linguistic ambiguity, to retrieve a red nozzle before oxygen levels on their ship depleted (Breazeal, et al. 2013). Two conditions of this were tested to examine the robustness of human aversion of risk to either a robot or another human given a robot with a narrative backstory versus one without — representation of the human teammate was kept constant. Results pointed at a possible threshold for risk averse versus risk seeking decision strategies based on increased social traits in the robotic teammate.

I. Introduction

Risk aversion is a decision theory concept that occurs when a decision maker knows the resultant outcome but is unaware of the outcome probabilities; in turn, she will more likely that not side with the choice where there is a subjectively known outcome probability versus one in which the probability outcome is unknown. The established model of this concept is known as the Ellsberg Paradox — often cited as the repudiation of the normative decision risk theory, Expected Utility (EUT), and evidence for distinction between unknowable *ambiguity* and uncertainty versus cognitively *computable* risk (Coleman, 2011). In contexts of real-world decision-making, the Ellsberg paradox holds credence because real life decisions are routinely ambiguous in expected outcome: i.e. which car to buy? Should you side with Alphabet vs. Apple stock in Q4? Is Donald Trump pro or anti-climate change? In most cases, decision makers must subjectively extrapolate their outcome estimations in the face of absent, or ill-defined probabilities — a gamble most of us subconsciously make every day. These decision models have been studied extensively in intra-human and risky choice contexts where subjects are given a list of options — with some more ambiguous than others by linguistic or visual means — then asked their preference. But what about situations in which humans are functionally augmented via the assistance of autonomous agents? Do our attitudes of risk aversion and preferential treatment of more defined outcomes hold in these situations where robots have the ability to make decisions for us?

This paper presents a study uncovering where humans defer risk in HRI collaborative work given the social affordances of robotic teammates. Risk here is in ambiguity of outcomes and a gamification aspect of depleting oxygen levels on a spaceship. This decision stress test was designed to test the relationship between robot utility values and human decision-making in HRI via replication of the "Ellsberg Urn" ambiguity design (Pulford, Coleman, 2008). The study was conducted on Mechanical Turk, which offers a much more diverse sampling of interaction types that can be transferred to physically embodied HRI systems. Furthermore, research done by Cynthia Breazeal has shown the utility of transferring online interactive learning and data-driven models as an alterative to hand gathered data techniques (Breazeal, et al. 2013). This is especially important when most robotics research fails to capture diversity in human-robot interactions, often training on subjective one-sided data sources. And as online interactions connect themselves to physical devices and autonomous systems in the real world, extracting these nuanced behavioral patterns in decision theory and risk aversion is a critical direction in the future of robust HRI research and systems design.

II. Background

A. *Robot Effects on Group HRI*

Psychology-based HRI research has shown that human treatments of robots in interactive settings are similar to how humans treat other humans in similar cases (Paepcke, Takayama, 2010). And like with human-human interaction, these treatments are dictated by social cues. Fraune et al. has shown that human interactions with social, anthropomorphic afforded robots were much more prolonged and positively rated than similar interactions with utility or mechanomprphic robots (Paepcke, Takayama, 2010). Similarly, human-to-robot affect perception has been shown to result in higher group sociality scores, potentially dictating the ways in which humans perceive the overall usefulness of a robot in a collaborative setting based on its social cues (Jung, dePalma, et al. 2012). These findings are an interesting point of research because they can explain how well traditional HHI (human-human interaction) behavioral models can transfer to HRI collaborative tasks depending on social affordances of the robot. Similar work done by Darling et al. demonstrates that empathetic narratives and characteristics such as backstories and names have an affect on HRI engagement types (Darling, Nandy, Breazeal, 2015). If this assumption holds, it could be hypothesized that risk aversion as it pertains to humans, would have a similar effect when robots are thrown into the mix.

I took inspiration from this literature in analyzing the two conditions in which the robotic teammate was given a socializing narrative or not. Keeping a humans' penchant to anthropomorphize human-like devices, ambiguity aversion via the Ellsberg decision-making model could provide a rich dataset for real-world applications of interaction models by which humans and robots navigate contexts where information is sparse and risk is high.

Research done exploring goal inference in human-robot collaboration shaped the level of "adaptivity" affordances I chose to give to the robot in my study design. Results from Liu, Hamrick, et al.'s research have shown that the *perceived* level of adaptivity in a robot — its social ability to mesh with the human worker instead of being a mechanamorphic tool — increased the perception of that robot to the human, improving the objective and perceived performance of the robot to the human worker (Figure 1) (Liu, Hamrick, et al. 2016).

B. *The Ellsberg Paradox & Its Experimental Utility*

As noted above, the task designed for this experiment was inspired by both the "Ellsberg Urn" ambiguity study, as well as the Mars Escape Human Commutation

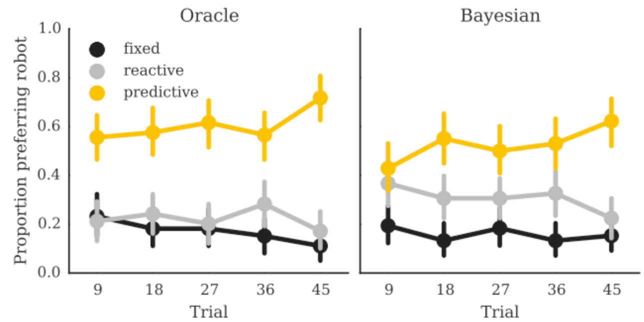

Figure 1: Oracle vs. Bayesian adaptivity trials. The yellow line indicates human preference increased as the robots' predictive traits increased (Liu, Hamrick, 2013)

game from Breazeal et al.'s Crowdsourcing Human Robotic Interaction study (Breazeal, et al. 2013), (Pulford, Colman, 2008). Ellsberg Paradox theory holds that when faced with computationally certain versus ambiguous outcome probabilities, humans are more likely to choose the certain.

In his 1961 illustration, Ellsberg had 2 Urns, each filled with red and black balls. Participants were told that in Urn A, there were approximately 50 red and 50 black balls, mixed randomly. In Urn B, there was an unknown ratio of 100 red and black balls mixed randomly. While blindfolded, participants would have to draw a red ball to win a financial compensation: which Urn would they choose to draw from?

The results showed that decision makers would overwhelmingly prefer Urn A for its more definite risk outcome, regardless of which color they were instructed to draw. This is known as the ambiguity aversion affect (Pulford, Colman, 2008). What makes this aversion affect such an interesting decision theory to test is that it is regarded as a suboptimal human decision making mechanism, violating the rational decision making model of subjective expected utility (SEU). SEU is the computational gold standard for evaluating decisions under uncertainty (Nau, 2007). It proposes that under uncertainty, there is no way to calculate relative probability, thus people will decide based on subjective calculations weighing the pros and cons of all summed probabilities (Nau, 2007).

In the case of the 2 Urns, an optimal human (imaging that there is such a thing) would have an equally likely outcome of choosing A or B without favoring one for its uncertainty. Other experiments have shown a statistically consistent favoring of finite over ambiguous outcome probabilities despite some of the more ambiguous ones

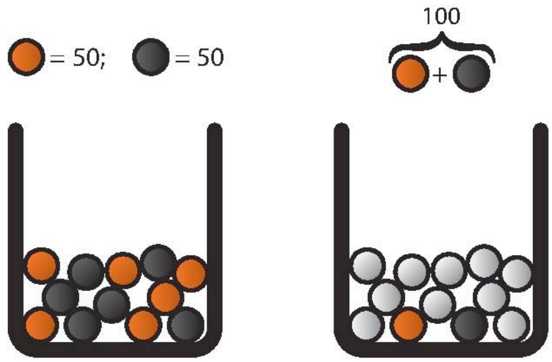

Figure 2: Ellsberg 2 Urn illustrating computable risk vs. unknowable ambiguity (Ellsberg, 1961)

actually having the same, if not higher chance of success than the finite ones.

There is a wealth of research done on the ambiguity aversion affect in human-human interaction (HHI) and decision making social psychology research. Yet little has been done testing the phenomena in collaborative group decision making contexts, yet alone HRI — and that's where my work differs. When you consider the affordances a socialized robot can bring to humans as augmenting partners in decision making, there may be a possibility to correct them against an ambiguity aversion bias, optimizing their decision making strategies in the face of risk.

C. *Crowdsourcing HRI*

Crowdsourcing data to leverage the wisdom of crowds allows for a collection of diverse interaction styles for training HRI data sets and testing theories in human interaction. On Mechanical Turk, researchers are provided with quick and translatable datasets that bolster traditional, "hand gathered" AI and machine learning methods by leveraging the power of the Internet via gamification and incentive-based work. Social psychology and decision theory research has long made use of crowdsourcing to test theories in perception, causal induction, probabilistic reasoning, and statistical models of language. Boi Faltings and Radu Jurca's study on incentives to counter bias in human computation made use of Mechanical Turk to test the robustness of peer truth serum and Bayesian truth serums on countering heuristic biases (Faltings et al., 2014). In research behind the cognitive process of human strategy selection, Lieder and Griffiths, formulated a theory that humans can gradually learn to make use of fallible decision making heuristics — testing 120 Mechanical Turk workers with 30 binary decisions to test strategy selection (Lieder, Griffiths, 2015).

The utility of a data-driven, crowdsourcing addendum to HRI research has been realized in recent research, as robots require a wide range of interaction types to enable seamless collaboration with humans. Cynthia Breazeal's research replicated the Restaurant Game online to extract human interaction patterns to train real-world robots on goal-oriented tasks (Jung, dePalma, Chernova, Hinds, Breazeal, 2013). Breazeal's group at the Media Lab has also extended HRI crowdsourcing to research in interactive systems between humans and robots around collaborative tasks and exploring mechanisms in robot learning.

A major issue with leveraging remote crowds-workers is in motivating them to do high quality work with little to no pay. Research done by Arpita Ghosh has looked at novel ways to incentivize workers via games with a purpose (GWP) (Ghosh et al. 2013). Ghosh's research, and the motivation behind why I chose to elicit a GWP model to test the ambiguity aversion affect in HRI collaborative work, stems from a growing literature in social psychology research that addresses what motivates people to behave and contribute quality work. GWP is a human computation system by which humans are playing a game, while at the same time, producing information around a task or computation that computers fail to replicate (Ghosh, 2013). The incentives behind this is to keep users engaged at a cognitive level — via reward schemes like points, badges, financial payouts, and rankings — so that they output useful data. In studying human decision making models, it is crucial that participants are cognitively engaged so as to control for outside decision making heuristics and distractions that could potentially skew results.

I enlisted a GWP replicate model of the Mars Escape game to situate crowds-workers into the role of a captain making decisions for her crew before oxygen ran out on their ship (Brezeal et al, 2013). Following the GWP motivation scheme of financial incentive, participants were given a financial scheme by room, dependent on which choice they made (this will be fleshed out in full in the methods section). This proximity was designed to simulate real-life decision making as if the risk were real, accurately testing ambiguity aversion.

Building on the above work, my research looks to test the Ellsberg ambiguity paradox as a decision-making strategy in the context of a GWP model. This approach was taken to better understand how a diverse set of humans averted ambiguity in HRI collaborative contexts. The results could potentially be used to extract models of decision making to apply to real-world contexts in which humans interact with autonomous agents — potentially augmenting our perceptions of ambiguity and making us better decision makers.

D. Overview

To test my hypotheses around how humans will embody risk aversion in an online HRI collaborative work, I conducted an experiment in which 2 random cohorts of 50 Mechanical Turk workers were assigned either condition 1: a robot with no socializing narrative (a tool), or condition 2: a robot with a socializing narrative (a companion). My hypotheses are as follows:

H1: For the condition where the robot is given <u>no narrative</u> (a functional tool), participants are more likely to send it to the room where there is an unknowable ambiguity

H2: For the condition where the robot <u>is given a narrative</u> (social research companion), participants are equally as likely to send it to the room where there is a computable risk as the human teammate

H3: For the human teammate in condition 1, participants are more likely to send it to the room where there is a computable risk over the robot

H4: For the human teammate in condition 2, participants are equally as likely to send it to the room where there is a commutable risk than where there is an unknowable ambiguity — they should view risk aversion equally between the robot and human

Based on Ellsberg's Paradox, humans are more likely to side with computable risk over ambiguity. Therefore, if the robot is given a narrative that makes it more "human," it too should be used to favor the computable risk (Ellsberg, 1961).

III. Methods

A. Experiment Design

To experiment using the paradox model in an HRI collaboration, participants were given a mission control scenario in which they — as the captain of 2 fictional teammates at a research base on Mars: 1 human, 1 robot — were directed to decide which teammates were to blindly retrieve the correct red nozzle from boxes placed in 2 rooms.

	Description
Room A	You recall there being a box of approximately 50 Red and 50 Blue nozzles
Room B	Your memory is a bit hazy, but you recall there being a box with an unknown ratio of 100 Red and Blue nozzles

Table 1: Room A and Room B conditions

	Conditions	
	1	***2***
Vrinda Harris	Exploration Engineer	Exploration Engineer based in Seattle. Vrinda has worked with you team on 2 previous missions and is super talkative
Robot	Exploration Engineering Bot	Oliver (Robot): He's an Exploration Engineering Bot from Pasadena CA. Oliver has been used by your team with 2 previous missions and is known to be very curious

Table 2: Narrative backstory for robot in Condition 1 and 2

The objective was to fix their rocket launcher before oxygen levels on their ship depleted. The task was timed at exactly 5 minutes. Diligent Turker's who read instructions would be left with approximately 1 minute to decide where both teammates went (see appendix A).

As stated, the purpose of this study was to test human aversion towards ambiguity when faced with risk, and how that was affected by social or non-social robots, thus there was no verifiably correct sequence for deciding which teammate went to which room. This was a replication of the Mars Escape game as described in section C (Jung, dePalma, Chernova, Hinds, Brezeal, 2013). My study differed from that design by introducing a layer of ambiguity to test Ellsberg Paradox, as well as a simplification of tasks that were given to the participants. This study was also entirely text based in lieu of the Mars Escape GUI simulations. Instead of extracting an interaction model across a series of join-collaborative tasks as with Brezeal et al., I was specifically interested in their decision strategies around risk via linguistic ambiguity — a mimicking of likely encounters of everyday risk in the real world. In turn, the task that the participants had to assign their two teammates replicated the Ellsberg 2 Urn design introduced in section B (Ellsberg, 1961).

There were two rooms on the spaceship where the participant (the captain) could recall there being 1 box each of an

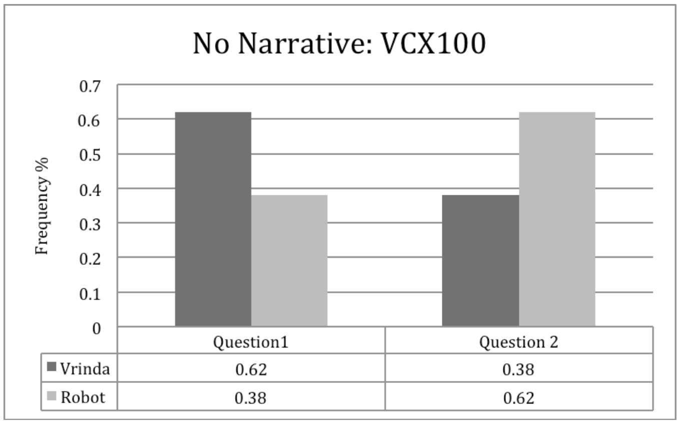

Table 3: Non Narrative frequencies of assigning human or robot to computable risk (Q1) and unknowable ambiguity (Q2) rooms

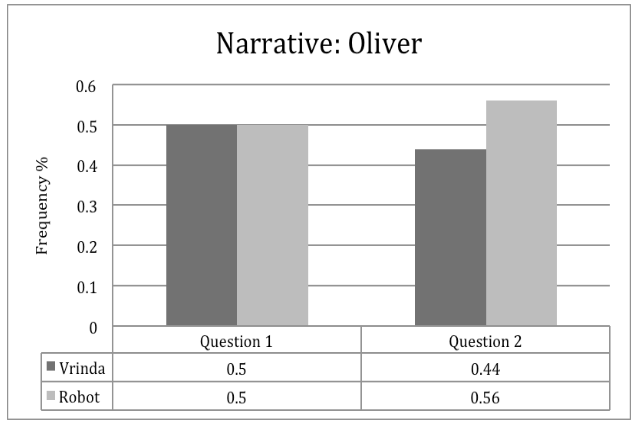

Table 4: With Narrative frequencies of assigning human vs. robot to computable risk rooms are equal.

assortment of red and blue nozzles. In Room 1, they could recall there being approximately 50 red and 50 blue nozzles. In Room 2, they were told that their memory was a bit hazy, but they could recall there being an assortment of *about* 100 red and blue nozzles combined (Table 1). The linguistic ambiguity of each box description was directly borrowed from Ellsberg's design in order to accurately test his decision theory model (Ellsberg, 2961)

Depending on the cohort conditions (1 or 2), participants were introduced to the respective description of their two teammates: a human named Vrinda Harris and a robot — both of which were described as being "Exploration Engineers" to keep their functional descriptions constant.

In *Condition 1* (non social), Vrinda was presented as: "Vrinda Harris: Exploration Engineer;" and the robot was presented as: "a VCX1000 Exploration Engineering Bot." In *Condition 2* (socialized), Vrinda was presented as: "Vrinda Harris. An Exploration Engineer based in Seattle. Vrinda has worked with your team on 2 previous missions and is super talkative;" and the robot was presented as: "Oliver (Robot): He's an Exploration Engineering Bot from Pasadena CA. Oliver has been used by your team with 2 previous missions and is known to be very curious" (Table 2) (appendix A).

B. Participants

I recruited participants on Amazon Mechanical Turk. A total of 100 workers participated in the study: 50 for HIT condition 1, and 50 for HIT condition 2. Workers who had already completed condition 1 were excluded via special qualification from doing the condition 2 HIT. Workers were compensated $.1 for the task, which was timed at 5 minutes. There was no special "master" qualification for workers.

C. Risk Logic

Before deciding which teammate to send to which room, workers were presented with a financially incentivized risk scheme (Ghosh, 2013):

Room 1:
If the drawn nozzle is Red: you win $.02 and can quickly fix your launcher before oxygen is depleted;

If Blue: won't work, you must defer to Room 2…

Room 2:
If this drawn nozzle is Red: you win $.01 and can fix your launcher with limited time before oxygen is depleted;

If Blue: won't work, you get $0 and your oxygen levels will run out before the launcher is fixed

Note that Room 1 and Room 2 would be simulated in succession and participants were told that no one teammate can do both rooms — i.e. if Vrinda was chosen for Room 1, the robot must do room 2, and vice versa. The other motive of using a financial incentive was to add a layer of risk to the scenario that oxygen levels were depleting and Turkers only had circa 1 minute to decide where each teammate went before the task timed out. This ultimately can aid in transferring the ambiguity simulation to real life scenarios where money is a studied risk scheme (Ghosh, 2013).

Once presented with their teammate descriptions, Room 1 and 2 conditions, as well as the risk scheme, participants were given two questions. Q1: in room 1 (given the description) which teammate would they assign? Q2: in room 2, which teammate would they now assign? In condition 1, they were presented with a selection of Vrinda Harris or the VCX1000 Robot, in condition 2 they were presented with a selection of Vrinda Harris or Oliver (robot).

IV. Results

In the first condition, in which the robot was not given a narrative, participants were more likely at .62 to assign the human to the computable risk room (approximately 50 red and 50 blue nozzles) than they were with the VCX1000 robot at a .4 probability. For the unknowable ambiguity (unknown ratio of 100 red and blue nozzles), participants were, on the reverse, least likely .4 to assign Vrinda than the robot — which they assigned with a frequency of .62. This upheld both H1 and H3 as participants were more likely to assign ambiguity to the robot, and more likely to assign computable risk to the human. (Table 3)

For the second condition where there was an added narrative to the robot, now known as Oliver, the results were mixed. For quality control, a few workers answered robot twice instead of one teammate per task. In this situation, I took the answer to question 1, room 1 as their preference and assigned the remaining teammate to question 2, room 2. In the narrative condition, participants were exactly equally likely to assign the human and the robot to the computable risk room: .25 and .25 respectively.

V. Discussion

I hypothesized that in transferring the Ellsberg Paradox to HRI, humans are still more likely to side with a subjectively known outcome probability versus one in which the probability outcome is unknown. In turn, when deciding weather to have another human or robot do these tasks, they will give the more computable outcome to the human. My results in this study do confirm this logic as it shows that human aversion to risk in ambiguity is robust enough that when deciding actors to do ambiguous tasks, they are increasingly more likely to pick non-human "tools" versus humans — despite that actor being a robotic *teammate*. When that robotic teammate is socialized by a behavioral traits and narrative, the relationship is neutralized, and humans are equally as likely, with no clear sign of distinction, to assign known probability outcome tasks to both teammates. This presents questions towards a deeper look into how we view robotic teammates.

In re-visiting the ambiguity aversion affect, as a decision strategy it's known as a fillable human bias. A substantial line of literature testing ambiguity aversion has consistently shown that even when two options are verifiably equally likely in their outcome probabilities, humans are still more likely to side with the less ambiguous option, even if that option isn't the best bet. While my experiment had two outcomes that were both non-verifiable in outcome, the result of humans still preferring less ambiguous options upholds the robustness of the paradox. But is this a bad thing? Is it bad that we prefer risky and ambiguous outcomes be dealt with non-human tools like robots? I argue not.

Human decision making strategies are sub-optimal and fallible to biases and incorrect heuristics that skew their outcomes. And while we can adopt our mental models and outcome simulations, we still fail in the face of risk and the unknown — especially when there is money involved. Research has shown this. Thus having a robotic partner, who likely won't be averse to ambiguity or biased by outside data, deal with risky decisions could prove to be an effective tool in correcting for human decision-making biases. So what does this mean for HRI design?

An interesting result from this experiment is that it again proved a known hypothesis that there is a threshold for how humans interact with robots depending on their level of social affordances. Oliver, the socialized robot in condition 2, was afforded the same ambiguity-averse caution as Vrinda from the Turker standpoint. What this potentially means is that in designing HRI tasks, the perceived utility versus socialization of a robot has an affect on how risk averse humans are. With more socialized robots, human risk aversion is normalized to a level where they are equally likely to attribute ambiguity and uncertainty on equal terms; they were more comfortable with having a robot do the higher payout, less riskier scenario. And this is likely due to trust via anthropomorphism. What would be an interesting perspective is how this relationship of trust changes the humans risk "appetite" on subsequent tasks. In condition 2, Vrinda was equally as likely to be assigned to the ambiguous task as Oliver the robot. This behavioral nuance is in itself a bias corrector, and it can be both a blessing and a curse. By lowering human guard to risk outcomes, successful leveraging can elicit more reckless decision strategies depending on the nature of the context. This was not the focus of my paper, but a growing body of literature addresses the affect of robots in advertisement and the ethical grounds by which they may nudge humans into siding with less secure outcomes (i.e. taking out a mortgage) (Darling, et al, 2015).

So a loose finding is that socialized robots may indeed make us less risk averse, and increase our tolerance of siding with ambiguous outcome probabilities — perhaps even to the point where we can correct for the Ellsberg Paradox's suboptimal tendencies. This decision theory model takes into account the affect that robots have on how humans make decisions — which is a crucial area of study as we advance towards a future of artificial intelligence and autonomous systems working along side us as companions. Little work has been done to transfer the wealth of decision theory research into the modern world, to plug robots and technology into the fold of human mental simulation models and probability outcome filtering. Understanding how they are situated as actants of human decision strategies in HRI (especially given the diversity of data from Mechanical Turk) can shape our understanding of robotic systems and HRI design models for future studies.

One of the primary limitations of this study was in a lack of due diligence to ensure that what was being measured wasn't subjective data on how risk averse or risk seeing individual participants were — this is partially due to the nature of Mechanical Turk and scope of this experiment. Past research on decision heuristics correct for subject risk appetite by applying peer and Bayesian truth serums to datasets. For my replication of the Mars Escape game, there was no statistically verifiable payout for each room. The probability outcomes for both were unknown. Therefore, confounding data could arise where more risk seeking workers were daring enough to have the VCX1000 robot take the calculable probability outcome, and more risk averse workers would prefer a human take it. This would have been measuring independent appetites for risk. Future studies using this design could filter participants based on their level of risk aversion by simply asking them at the beginning how risk averse or risk seeking they are. This would create some interesting findings that might anchor their choices given how they answered, or give us threshold measurements on how to nudge risk averse to risk seeking, and risk seeking to risk averse.

Another limitation was in the design of the task itself. Keynesian economic theory holds that: "If two probabilities are equal in degree, ought we, in choosing our course of action, to prefer that one which is based on a greater body of knowledge?" (Chew, Epstein, Zhong, 2007) (Keynes, 1921). Research in this area has proposed that decision makers depend not just on knowable probabilities, but also on how the uncertainty itself may arise (Chew, et al. 2007). Thus, in the context of my Mars Escape game, Turkers likely were not familiar with a scenario where their spaceship was running out of oxygen. In turn, that context alone could have had an affect on how likely they were to have a human take a safer outcome probability than a robot. Future research could look into testing ambiguity aversion on Mechanical Turk using more realistic decision-making contexts.

I also would have liked to see more iterations of this in order to test how robust the Ellsberg Paradox is. Is there a way to compute risk aversion? Is there a financial threshold by which participants suddenly become risk seeking — correcting for the paradox with cash? These are tantalizingly interesting questions that some future research could study. The $.02 and $.01 payment scheme was meant to add risk to their decision. But one can imagine that these along could actually minimize uncertainty by making people more risk seeking if the payout is high enough. Furthermore, if the incentive scheme was flipped, with a higher payout for success on the ambiguous outcome box, a natural tension could have been created, further testing the parameters of the ambiguity aversion affect.

VI. Conclusion

This experiment was meant to be an introduction into the space of decision theory in HRI — as my broader research looks at understanding the roles robots [technology] play in human decision-making and ways to correct for bias. The objective of this study was to test where humans deferred risk in HRI collaborative work given the social affordances of robotic teammates — under the context of risk via ambiguity. The more socialized the robot was, the higher chance humans would defer risk to both the robot and human teammate equally. The less socialized, the more likely the human would defer the unknowable outcome to the robot, averting risk for the human teammate. My findings upheld the paradox as transferring into online spaces and HRI, as well as upheld my hypotheses. However, what was also taken from the results is that socialized robots may themselves be bias correctors, making humans less risk averse depending on context. This has interesting implications in the way we design HRI collaborative systems by understanding a robotic teammate's affect on our decision-making strategy.

References

[1] Bombardini, Matilde, and Francesco Trebbi. "Risk Aversion And Expected Utility Theory: An Experiment With Large And Small Stakes." *Journal of the European Economic Association* 10.6 (2012): 1348-399. Web.

[2] Breazeal, Cynthia, Nick Depalma, Jeff Orkin, Sonia Chernova, and Malte Jung. "Crowdsourcing Human-Robot Interaction: New Methods and System Evaluation in a Public Environment." *Journal of Human-Robot Interaction* 2.1 (2013): 82-111. Web.

[3] Chang, Liu, Jessica Hamrick, Jamie Fisac, Anca Dragen, Karl Hedrick, Shankar Sastry, and Thomas Griffiths. "Goal Inference Improves Objective and Perceived Performance in Human-Robot Collaboration°." *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems* (2016): n. pag. Print.

[4] Chernova, Sonia, Nick Depalma, Elisabeth Morant, and Cynthia Breazeal. "Crowdsourcing Human-robot Interaction: Application from Virtual to Physical Worlds." *2011 Ro-Man* (2011): n. pag. Web.

[5] Chew, Soo Hong, Richard P. Ebstein, and Songfa Zhong. "Ambiguity Aversion and Familiarity Bias: Evidence from Behavioral and Gene Association Studies." *Journal of Risk and Uncertainty* 44.1 (2011): 1-18. Web.

[6] Coleman, Thomas. "Probability, Expected Utility, and the Ellsberg Paradox." *SSRN Electronic Journal* (n.d.): n. pag. *Close Mountain Advisors LLC*. Web.

[7] Darling, Kate, Palash Nandy, and Cynthia Breazeal. "Empathic Concern and the Effect of Stories in Human-robot Interaction." *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015): n. pag. Web.

[8] Joosse, Michiel, Vanessa Evers, and Lahose Manja. "Crowdsourcing Culture in HRI: Lessons Learned From Quantitative and Qualitative Data Collections." *Research Gate* (2016): n. pag. Print.

[9] Jung, Malte, Nick DePalma, Sonia Chernova, Pamela J. Hinds, and Cynthia Breazeal. "Human-Robot Collaboration: Bids and Bytes." *Journal of Human-Robot Interaction* (2012): n. pag. Print.

[10] Nau, Robert. "Extensions of the Subjective Expected Utility Model." *Advances in Decision Analysis: From*

Foundations to Applications (2007): n. pag. *Cambridge University Press*. Web.

[11] Paepcke, Steffi, and Leila Takayama. "Judging a Bot by Its Cover." *Proceeding of the 5th ACM/IEEE International Conference on Human-robot Interaction - HRI '10* (2010): n. pag. Web.

[12] Pulford, Briony D., and Andrew M. Colman. "Size Doesn't Really Matter: Ambiguity Aversion in Ellsberg Urns with Few Balls." *Experimental Psychology*. U.S. National Library of Medicine, 2008. Web.

[13] Toris, Russell, David Kent, and Sonia Chernova. "The Robot Management System: A Framework for Conducting Human-Robot Interaction Studies Through Crowdsourcing." *Journal of Human-Robot Interaction* 3.2 (2014): 25. Web.

[14] Weber, Bethany, and Wah Pheow Tan. "Ambiguity Aversion in Risk and Delay." *PsycEXTRA Dataset* (n.d.): n. pag. Web.

[15] Whysel, Brett. "Can Robots Save Us?" *Decision Fish*. N.p., 20 Oct. 2015. Web.

.

Appendix A

Instructions

Mission Control:

You and your team of 2 other scientist are preparing to leave your research base on Mars to travel back to earth, but there's one problem: the launch system on your ORION spaceship was damaged in a storm and needs to be repaired before the oxygen level on your ship is depleated (you've overstayed your reserve supply).

In order for the launch system to be repaired, and with time running out, you need 1 of your teammates to blindly (you've cut electricity to the ship to preserve power) draw **1 Red nozzle** out of a box of Red and Blue nozzles — but each teammate only gets one chance to draw the Red nozzle because opening a box once contaminates the other nozzles.

There are **2 rooms** on your ship that you know for certain contain **1 box** of nozzles each — you'll see descriptions of each room below. Your task is to decide which of your following teammates goes to which room before time runs out. No one teammate can do both rooms, and neither can see because you've cut power.

Your Teammates:

- Vrinda Harris: Exploration Engineer based in Seattle. Vrinda has worked with your team on 2 previous missions and is super talkative
- Oliver(Robot): He's an Exploration Engineering Bot from Pasadena CA. Oliver has been used by your team with 2 previous missions and is known to be very curious

The Conditions:

Room 1: If the drawn nozzle is Red: you win $.02 and can quickly fix your launcher before oxygen is depleted

If Blue: won't work, you must defer to Room 2...

Room 2: If this drawn nozzle is Red, you win $.01 and can fix your launcher with limited time before oxygen is depleted

If Blue: wont work, you get $0 and your oxygen levels will run out before the launcher is fixed

You're the captain, now decide who goes where!

***Remember:** *no one teammate can do both rooms (you don't have time), so your first choice can't also be your second*

Room 1:

Description: you recall there being a box of approximately 50 Red nozzles and 50 Blue nozzles.

Which teammate do you assign here?

- ○ Vrinda Harris
- ○ Oliver (Robot)

Room 2:

Description: your memory is a bit hazy, but you recall there being a box with an unknown ratio of 100 Red and Blue nozzles.

Which teammate do you now assign?

- ○ Vrinda Harris
- ○ Oliver (Robot)