Influence of Goal Alignment on Delegation Decisions to Human and Automated Collaborators

by

Jessica Lee

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved February 2024 by the
Graduate Supervisory Committee:

Erin K. Chiou, Chair
Nancy J. Cooke
Heather C. Lum

ARIZONA STATE UNIVERSITY

May 2024

ABSTRACT

Automation is becoming more autonomous, and the application of automation as a collaborator continues to be explored. A major restriction to automation's application as a collaborator is that people often hold inaccurate expectations of their automated collaborator. Goal alignment has been shown to be beneficial in collaborations and delegation decisions among human-human and human-automation collaborations. Few studies have investigated the difference that goal alignment has on human collaborators compared to automated collaborators. In this 2 (goal aligned or misaligned) x 2 (human or automated) between-subjects study, participants complete a simplified triage patient task and then are given the opportunity to stay with their manual task solution or to delegate their decision and go with their collaborator's recommendation. Participants never delegated to collaborators with goals that were not aligned to theirs. Participants working with human collaborators that have similar goals to them were more often delegated to and more often associated with a better triage performance. These results can inform the design of similar systems that foster collaboration and achieve better team performance. Although goal alignment was crucial for delegation decisions, it was difficult to achieve complete agreement of goals. Future research should investigate effective methods to better communicate goals among collaborators.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Complex work systems require complex solutions. These complex solutions can help numerous fields such as the military, autonomous vehicles, and healthcare. All of these fields require complex decision-making and the productivity of the systems relies on the coordination of people from multidisciplinary backgrounds effectively working together (Pype et al., 2018; Xiao & Mackenzie, 1998) However, the people within these complex systems are often overloaded with decisions to be made and, because complex systems are dynamic (Sterman, 1994), some people within the system may find themselves with extra capacity when another person is overloaded. To better balance the workload among everyone in a complex work system, people can delegate tasks to one another (Richards & Stedmon, 2016).

Healthcare is a complex system that needs a way to offload decision-making. Additionally, because healthcare workers share a common goal of providing high-quality care to the patient, they are more likely to seek expertise from their peers and collaborate with one another (Pype et al., 2018). Healthcare workers are often compelled to delegate because they operate with insufficient resources (Knickman & Snell, 2002; Walker & Gilson, 2004) in a high-risk environment which makes their time and energy to make complex moral decisions limited (Dzindolet et al., 2002; Liehner et al., 2022; Lyons & Stokes, 2012). When healthcare workers become overworked, they routinely delegate tasks to one another as a way to cope with the overwhelming workload (O'Malley et al., 2015; Ridde et al., 2012; Yukl & Fu, 1999). However, with the shortage of healthcare workers (Leong et al., 2021), fewer people have the capacity to take over these complex decisions.

Automation has been considered a potentially useful tool to add to these complex systems, such as healthcare because automation can assist people in their tasks (Miller & Parasuraman, 2007; Sun & Botev, 2021). In particular, automation that acts as a decision support system can be helpful in offloading tasks from people. Because automation is typically an assistant to people, they are less likely to be overwhelmed with decisions and often have the capacity to receive delegated decisions (Parasuraman & Riley, 1997).

However, it is not easy to introduce a collaborator (automated or otherwise) into a new environment and expect it to immediately acclimate to the healthcare environment. The decisions that healthcare workers routinely make are cumbersome and complex, balancing many factors. A collaborator must be able to independently perform well enough so that the worker can trust it and be willing to collaborate and ultimately delegate decisions to that collaborator (Lee & See, 2004).

Background Literature

To best implement and facilitate these collaborators, it is crucial to understand what makes people willing to collaborate and ultimately delegate their decisions to others. Part of unpacking this reasoning involves understanding the involved collaborators, and what collaboration entails. Collaboration occurs between two agents interacting and "behaving together, in some relation to one another…who also have some past and/or future relation to each other" (McGrath, 1984). Collaborative agents can possess any number of goals that are not necessarily shared among the other agents it is collaborating with, so calling this collection of collaborative agents a "team" might be misleading (Salas et al., 1992). Additionally, collaborative agents can take many physical forms. Sometimes these collaborators can take the physical form of other people, but more recently, there has been an influx of automated collaborators bearing delegated tasks. Automated collaborators are thought to be a simple solution to alleviating the task overload of workers (Miller & Parasuraman, 2007; Sun & Botev, 2021). Collaborators could take many forms (e.g., a pencil, a chair, a pet), but people expect that other people and automation will have an exceptionally high capability of making a good decision (Dijkstra, 1999; Dijkstra et al., 1998; Dzindolet et al., 2002). Decisions can only be appropriately delegated to those that could perform the task on behalf of the person (Cheong, 1996), therefore, people and automation are the two forms of primary focus of potential collaborators.

However, just because people and automation are perceived as the most capable collaborators does not make them equal. When people collaborate with other people, they expect their collaborator to be skilled at adjusting to unprecedented circumstances (Bainbridge, 1983; Sheridan, 1995) and would typically prefer to collaborate with another person (Dzindolet et al., 2002) for these unprecedented situations. On the other hand, people collaborating with automation expect their collaborator to be highly skilled at parsing through information quickly and operating efficiently (Bainbridge, 1983). For tasks that are believed to be better suited for these strengths of automation, a person will have higher expectations for an automated collaborator's performance than they would for a human collaborator's performance (Dijkstra, 1999; Dijkstra et al., 1998; Dzindolet et al., 2002; Lyons & Stokes, 2012). Similarly, when people are paired with a human collaborator for an automated-preferred task, they may be more likely to assume their human collaborator will perform inconsistently (Madhavan & Wiegmann, 2007) and pay extra attention to their human collaborator's contribution. This extra attention can lead to improved performance (Wickens et al., 2023). However, this potentially improved performance only occurs because people do not trust their human collaborator's performance for that particular task. Because of the different expectations people hold for each form of collaborator, the willingness to collaborate with each form may vary highly depending on the situation.

Yet sometimes, effective collaboration is still inhibited because the perceived strengths and expectations for each form of collaborator are misguided and result in inappropriate willingness to collaborate. Misguided expectations often arise from under-trusting a collaborator. When a person trusts a collaborator less than they should, they will be less willing to collaborate and therefore under-use their collaborator, resulting in less willingness to delegate their decision. (Dzindolet et al., 2002; Lewandowsky et al., 2000; Muir, 1994; Pruitt & Kimmel, 1977).

It is important to understand and bolster people's willingness to collaborate so that they are eventually willing to delegate decisions appropriately. When collaborating with other people, people expect other people to perform inconsistently (Madhavan & Wiegmann, 2007). This expected inconsistent performance lowers the trust a person has in their human counterpart and lowers their reliance on their human collaborator (Lyons & Stokes, 2012). However, the reduced trust may result in a more careful verification of their human collaborator's work (Patterson et al., 2007). When collaborating with automation, people expect automation to perform very consistently, and almost even perfectly. People expect automation to perform so well, that any sight of error, their trust in automation, and its ability to perform well, disproportionality drops (Dzindolet et al., 2002; Madhavan & Wiegmann, 2007). When the expectations held for a collaborator are hastily assumed, people are less likely to appropriately trust their collaborator – human or automated. Therefore, there is a need to establish trust of a collaborator more accurately.

A common argument as to why trust is easily inappropriately allocated is due to a lack of transparency of the collaborator's overarching goals. Without understanding the goals behind a collaborator's actions, it is difficult to predict the collaborator's actions in the future and trust is inhibited (Rempel et al., 1985; Trzebiński & Marciniak, 2022). Past studies have tried to investigate how human and automated collaborator's goals could be made clearer and their positive impacts. Most of these studies regarding clearer goals in human-automation collaborations are theoretical or anecdotal studies that state the importance of goal alignment, rather than empirical (Hall, 2003; Klein et al., 2004; Li & Lee, 2022). Most empirical studies that investigate human-automation collaborations are focused on aligning outcomes and performance (Chiou & Lee, 2015; Lewandowsky et al., 2000; Liehner et al., 2022; Lyons & Stokes, 2012) instead of goals. There is a severe lack of empirical research that demonstrates the influence of goal alignment in human-automation collaborations on a person's trust, expectations, and treatment of an automated collaborator.

Shared goals can reestablish expectations

Sharing goals can be beneficial because it allows people to better guess the intent behind each collaborator. A goal is an objective or aim of an action (Locke & Latham, 2012). Goals can often be aligned and shared among those in a group. Shared goals between human collaborators have been overwhelmingly associated with positive impacts. Group goals foster a willingness to engage (Mannix & Neale, 2005) and cooperate with each other (Bogaert et al., 2008). Sharing similar goals is so influential that even when people who did not want to collaborate were assigned a shared group goal, everyone cooperated more than groups without a shared goal (Bostyn et al., 2023). The shared goal allowed people to establish the expectations of their collaborators more accurately and led to a more appropriate willingness to collaborate with each other. This ability to join people's efforts together through better-established expectations makes goal alignment potentially very influential in fostering trust and cooperation between collaborators beyond just human-human collaborations.

Goal alignment has been predominately studied among human-human collaborations, typically in the work organization fields (Foddy et al., 2009; Krebs, 1975). Many studies have been theoretical studies, discussing the concept of optimal collaboration between people. When collaborating with other people, people who share similar personality traits and ideologies prefer to collaborate with one another (Bogaert et al., 2008). The same preference holds true with behavior – people who behave similarly tend to want to collaborate with one another (Becchio et al., 2010; Bostyn et al., 2023; Fitzsimons & Lehmann, 2004). Goal alignment is influential in collaborations between people. Many studies have investigated goal alignment in human collaboration using trust-related games in which one person can make a decision that would put them at a disadvantage unless their collaborator made the same decision. In these studies, the participants tended to favor collaborators who demonstrated the same behavior as themselves, and disfavored collaborators who were dissimilar to themselves (Becchio et al., 2008; Bostyn et al., 2023; Krebs, 1975).

The goal alignment benefits from collaborating among humans would likely similarly translate to collaborating with automation. Because people cooperate with automation in a similar way to how they cooperate with other people (Lewandowsky et al., 2000; Nass et al., 1996), it is plausible to assume that goal alignment would be similarly influential in human-automation interactions as it is in human-human interactions. This idea has already been indirectly supported through past studies. Past studies have shown that when people can customize automated products and services so that they better align with their needs and desires (e.g., picking which information is relevant to be displayed), people are more trusting of that automation (Koufaris & Hampton-Sosa, 2004; Shao et al., 2019; Sia et al., 2009). Additionally, revealing what the automation would do in a decision-making situation and the reasoning behind their potential decision led to increased trust, acceptance, and assistance in reaching the participant's goals (Verberne et al., 2012). From these past studies, it has been demonstrated that when the person understands and expects that the automation is working favorably with their own goals, willingness to collaborate increases, and delegating their decision to the automation increases.

Although past studies have established that people's willingness to collaborate is influenced by goals (and the establishment of expectations that come with these goals) for both human and automated collaborators, many of these studies end with the participant's action to collaborate or not, only testing with human collaborators or automated collaborators. Very little research has been conducted that focuses on the comparison of collaborator form (human versus automated collaborators) in combination with goal alignment status, and how those combinations affect delegation decisions.

Additionally, very few studies have extended their assessment to include the performance that results from the participant's decision to collaborate or not. If the performance is better when collaboration is rejected, even if both collaborators share the same goal, then collaboration may not be as beneficial as past work assumes. Conversely, performance may vary by the decision to collaborate (or not collaborate) when paired with an automated or human collaborator.

Current Study

People often lack appropriate expectations of their collaborators, which leads them to under-trust their collaborators, whether human or automated (Dzindolet et al., 2002; Madhavan & Wiegmann, 2007). This under-trust compels people to under-utilize their collaborators and reap the benefits of collaboration. To re-establish the expectations of collaborators and make the collaborators more attractive to delegate to, this study proposes the benefit of shared group goals (Bogaert et al., 2008; Bostyn et al., 2023). These shared goals have encouraged participants to work together and strive towards a common outcome with their collaborator, regardless of whether the collaborator is human or automated.

This study aims to investigate the influence of goal alignment among human and automated collaborators on a person's delegation decision. Additionally, this study investigates how participants' delegation decisions affected decision-quality performance based on a pre-determined point system and objective optimal solution. The four hypotheses for this study are below:

H1: People working with goal-aligned collaborators will delegate decisions more often compared to the people working with goal-misaligned collaborators.

H2: People working with automation will delegate decisions more often compared to the people working with other people.

H3: People working with goal-aligned automation will delegate decisions the most often compared to the other groups in the study.

H4: People working with human collaborators will have the highest decision-quality.

METHOD

To examine the decision-making collaborator characteristics that influence delegation techniques and performance, this experiment used a 2 (goal alignment status: aligned vs. misaligned) x 2 (collaborator physical form: human vs. automation) between-subjects factorial design. Participants were randomly assigned to one of the four conditions: paired with a human with goals aligned (H-GA) collaborator, human with goals misaligned (H-GM) collaborator, automation with goals aligned (A-GA) collaborator, or automation with goals misaligned (A-GM) collaborator. Because delegation likelihood varies heavily by the context in which the decision must be made (e.g., high-risk, moral dilemma, Dzindolet et al., 2002; Liehner et al., 2022; Lyons & Stokes, 2012), the task was designed to engage moral judgments in a somewhat risky environment while also incorporating objective numerical optimization. This context was selected to address the most consequential situation in which delegation could be highly beneficial or highly detrimental. Delegation performance was measured by the quality of their decision. The quality of the participant's decision was measured by their final decision compared to the most optimal solution and how well they adhered to given restraints. Open-ended questions were asked following their finalized output to further understand participants' goals and their decision-making process.

Participants

An a priori power analysis with a medium effect size (d = 0.25), power of 80%, and alpha of 0.05 indicates the need for 124-128 participants. Participants were required to be at least 18 years old, have access to a computer or laptop with Zoom installed, and be fluent in English. Participants who did not meet these requirements were excluded. Participants could be compensated with either 1 course credit hour or could be entered in a drawing for five $25-worth gift cards. In total, 133 participants were recruited, primarily from a large southwestern university. One participant was removed because they did not provide coherent responses to the open-ended questions. This resulted in 132 participants.

Task Environment

Participants were told that there would be 20 potential patients. However, the participant could only treat 13 of those 20 patients. The participant's task was to determine which 13 patients should be treated. Each of the 20 patients had associated profit, relationship, and severity points. These points estimate the amount of money earned from treating that patient, how much positive investment treating that patient would generate (i.e., the one-on-one time between the patient and physician alluding to the quality of care the clinic can provide), and the severity of the patient's condition, respectively. When choosing their patients, participants were instructed to maximize each set of points and achieve the highest overall points. Additionally, the second goal participants were instructed to adopt was adhering to a point differential target which will be described later.

Procedure

Participants arrived in the virtual meeting room and were sent a link to the consent form. Once participants gave their informed consent, participants were told to imagine themselves as a primary care physician. Their task was to determine which patients should be treated at their clinic.

First, the experimenter went over a tutorial patient selection task in which participants were taught the essential skills to select patients. The tutorial Google Sheet had 20 potential patients. In addition to understanding how to read each patient's points, the participants were instructed on how to select patients and how to read the automatically generated point sums for the patients selected.

Once the tutorial was complete, the participants were sent their own Google Sheet like the one used in the tutorial, but with a new set of 20 patients. All participants received the same set of 20 patients for the experimental task. Participants were instructed to select which 13 patients they would like to treat among the 20 presented, taking as long as they needed to decide. Right before sending participants their experimental Google Sheet, they were reminded of the two goals they are instructed to adopt. Participants verbally stated when they were finished making their selection. Once they stated they were complete, participants could not change their selection.

After the participant's manual selection, the participant was introduced to their randomly assigned one of four decision-making collaborators. Participants were told that their collaborator had picked their own 13 patients from the same set of 20 patients which produced the summation of their patients' profit, relationship, severity, and overall points (same generated sums that participants had access to for their manual selection). These decision-making collaborators were not real, and their answers were predetermined. Participants were shown the four sums from the collaborator's selected 13 patients, but not which 13 patients were selected. The participant could choose to accept the collaborator's recommendation (e.g., delegate their decision) or reject its recommendation completely and keep their self-generated patient selection. Participants could take as long as they needed to decide between their own manually generated solution or to delegate to their collaborator's generated solution.

Once the participant verbally said their delegation decision (manual solution or collaborator's solution), the experimenter conducted a semi-structured interview with the participants to understand why they ultimately accepted or rejected the collaborator's recommendation and what decision-making process was used. After the questioning was complete, the participant filled out a demographics survey, was debriefed, and the study concluded.

Materials

Decision-making collaborators

Participants were randomly assigned to only one of four conditions: H-GA collaborator, H-GM collaborator, A-GA collaborator, and A-GM collaborator. Two of these conditions involved goal aligned collaborators, and the other two conditions involved goal misaligned collaborators. All four conditions had an equal overall sum of points, but the collaborators with goals aligned met the point differential target, whereas the collaborators with goals misaligned did not meet the point differential target. The collaborators with aligned goals recommended the maximum number of points possible while adhering to the target discussed in the next section (the optimal decision-quality). The collaborators with misaligned goals recommended the maximum number of overall points but did not adhere to the guidelines. The collaborators with goals misaligned did not adhere to the guidelines by prioritizing the wrong set of points than what is instructed by the system. In this way, the goal aligned collaborators met both two goals, and the goal misaligned collaborators only met one of the two goals.

Each of the four collaborators had a three-sentence background description including their physical form, past work application, and how the collaborator typically prioritizes profit and relationship among patients. Each collaborator's description and their recommended points can be found in Appendix A.

Point system

Each of the potential 20 patients that the participant could treat was accompanied by their profit points, relationship points, and severity points, as described earlier. Each set of points was given on a 1-5 scale where 5 represents the most monetary profit, the most positive relationship investment, or the most severe condition, respectively. These three sets of points highly simplify common factors that are important when making a triage decision. The points for each patient were randomly generated.

To encourage participants to prioritize relationship points, a point differential target guideline was implemented. Participants were told there was a recent new guideline from the primary care physician union that states their total relationship points should be at least 10 points higher than the total profit points. If the point differential target was not met, the participant would incur a penalty. For every point that the total relationship points are less than the total profit points, the participant's clinic would lose 10% of its client base. Therefore, if the total relationship points and total profit points were equal (or total profit points were greater than total relationship points), the participant's clinic would lose 100% of its client base.

The goals outlined were to meet the 10-point differential target and to maximize the overall sum of points. Mathematically, this would result in the two goals-aligned collaborators' recommendation. All goals, however, will likely try to maximize severity points to treat those that most need care as a secondary goal.

Measures

In addition to the delegation decision, there were additional measures recorded. These measures will be discussed below.

Decision-quality

Decision-quality was measured from the participant's total sum, deducting proportionally to the magnitude to which they fell short of the 10-point guideline. The optimal solution is the 13 patients that reach the maximum amount of points possible while adhering to the 10-point guideline between relationship and profit points. If the participant's total sum (either their manually generated solution or the accepted collaborator's solution) did not meet the 10-point guideline, the participant's total sum was proportionally scaled down based on the extent to which it did not meet the guideline. For example, if the participant's raw score was 150 but they only had a 9-point difference between relationship and profit instead of the specified 10, their total sum would be reduced by 10% to achieve their final score of 135 (150 x 0.9 = 135). However, if the 10-point guideline was met, there would be no deduction because the total sum would just be multiplied by 1.0 to reach an identical final score. This final score is considered the decision-quality. This decision-quality represents the performance of the participant's final decision.

Semi-structured interview

Once the participant had made their delegation decision, the experimenter asked the participant general questions about their decision-making process. From these questions, the experimenter may follow up with clarification questions or questions to understand their values (e.g., "what was your priority of points?"). Sometimes, participants volunteered information about hypothetical situations (e.g., if the collaborator recommended different numbers).

RESULTS

Multiple chi-squared tests were conducted to test the four hypotheses and understand how goal alignment and physical form influence delegation decisions and quality.

To assess collaborator's goal alignment status and a participant's likelihood to delegate their decision to their collaborator and accept the collaborator's recommendation, a chi-squared test was run. A significant difference was revealed, [χ^2(1, N = 2) = 39, p = 0.4.238e-10, 1.0 = פ] (H1). This effect was clearly seen by zero delegations to collaborators with misaligned goals. About half of the participants paired with a goals aligned collaborator delegated to their collaborator, whereas the other half stayed with their manual solution. (Table 1).

Table 1

Count of Participants' Delegation Decisions Based on Goal Alignment

Collaborator's Goal Alignment and Delegation Decision	**Count**
Goals Aligned, Collaborator's Choices*	39
Goals Aligned, Manual Choice	27
Goals Misaligned, Collaborator's Choices*	0
Goals Misaligned, Manual Choices	66

*Two numbers compared in Chi-squared test for H1.

To test physical form's impact on delegating to a collaborator's recommendation, a chi-squared test was run, which revealed no significant difference between the likelihood to delegate to the human versus automated collaborator, [χ^2(1, N = 2) = 2.077, p = 0.150, 0.23= פ] (H2). 24 (36%) of the participants paired with a human collaborator delegated their decision to their collaborator. 15 (23%) of the participants paired with an automated collaborator delegated to their collaborator (see Table 2).

Combining goal alignment and physical form, the decisions to delegate were looked at across the four collaborators holistically. The number of participants that delegated their decision to their collaborator for each of the four collaborators is shown in the graph below (Table 2).

Table 2

Count of Participants' Delegation Decisions Based on Collaborator Type

Collaborator Type	**Number of Participants That Delegated**
H-GA	24
A-GA	15
H-GM	0
A-GM	0

The results of the chi-squared test found that there was a significant difference between the four groups, [$\chi^2(3, N = 4) = 43.154$, $p = 2.283\text{e-}09$, 0.61 = פ] (H3).

Finally, to investigate the characteristics of collaborators that result in the highest decision-quality, a chi-squared test was used. Below are the participants that resulted in the optimal solution and their assigned collaborator (Table 3).

Table 3

Count of Participants That Ended with the Optimal Solution

Collaborator Type	**Count**
H-GA	26
A-GA	18
H-GM	6
A-GM	2

There was a significant difference between the four collaborator types and their likelihood to end with the optimal solution, [$\chi^2(3, N = 4) = 28$, $p = 3.362e\text{-}06$, 0.42= פ] (H4). There was not, however, a significant difference between the number of participants that ended with an optimal solution between the human and automated collaborators across goal alignment statuses, [$\chi^2(1, N = 2) = 2.769$, $p = 0.096$, 0.18= פ] or between just the human-aligned and automated-aligned collaborators, [$\chi^2(1, N = 2) = 1.455$, $p = 0.228$, 0.23= פ]. It should be noted that just because a participant ended with an optimal solution does not necessarily mean that they delegated to their collaborator. 18 participants manually achieved the optimal solution (the same recommendation as a collaborator with goals aligned), which will be discussed next.

Exploratory Analysis

It was surprising that only about half of participants with goal aligned collaborators delegated to their collaborator. To further investigate this outcome, the 10 instances in which participants paired with a goal aligned collaborator manually achieved the optimal solution – a recommendation identical to their collaborator were examined. The delegation decisions and their paired collaborator can be seen in the table below (Table 4).

Table 4

Collaborator Type and Delegation Decision for Manually Optimal Solution

Collaborator Type and Delegation Decision	**Count**

A-GA, Manual	3
A-GA, Collaborator	1
H-GA, Manual	2
H-GA, Collaborator	4

The frequency of delegation decisions among those that manually achieved the optimal solution with the four goals-aligned collaborators was not significantly different, [$\chi^2(3, N = 4) = 2$, $p = 0.572$, 0.26= פ]. Additionally, there was no significant difference among the delegation decisions among the automated collaborators, [$\chi^2(1, N = 2) = 1$, $p = 0.317$, 0.50= פ], and among the human collaborators, [$\chi^2(1, N = 2) = 0.667$, $p = 0.414$, פ =0.33]. Although there was no significant difference observed, generally participants that manually achieved identical solutions to their collaborator tended to delegate to their human collaborator more often than to their automated collaborator.

In many cases, people delegated to their human collaborator because the participant felt that the collaborator had more expertise in the area than themselves. In other cases, participants delegated to their human collaborator to protect their collaborator’s feelings. Some participants even considered their collaborator’s future career promotions in their delegation decision to allow the collaborator an opportunity to showcase their work. Often, participants expected that being paired with a human collaborator would allow the opportunity to discuss each other’s reasoning. Participants rarely thought that it would be possible to converse with and understand the reasoning behind an automated collaborator’s recommendation.

To further examine participants’ delegation decisions when paired with goal aligned collaborators, it became clear from their open-ended responses that some participants held priorities that differed from their collaborators’ priorities. The goals-aligned collaborator was designed to match the goals imposed by the experiment design: maximize total points and to meet the point differential target. Nearly all participants met the point differential target to avoid penalties, but 42 out of the 132 (32%) adopted self-imposed goals. 38 of those self-imposed goals prioritized severity over relationship or total sum as the scenario outlined. In prioritizing severity, participants took an average loss of about 7 points from the optimal solution (146/153). Of the 93 manual decisions, 38 (41%) were from participants prioritizing the external severity goal over the total sum goal. Several of the participants that prioritized severity volunteered that they interpreted the task as a moral judgement in which severity should always be prioritized in any healthcare environment. Other, less common external goals adopted by participants include creating a point differential beyond the target, trying to balance all sets of points, and prioritizing profit points.

DISCUSSION

This study aimed to examine how a collaborator's goal alignment status and either human or automated physical form affects the participant's delegation decision and decision-quality. Among the four collaborator types (H-GA, H-GM, A-GA, A-GM), H-GA was most often delegated to and most frequently achieved the highest decision-quality. Although goal alignment was most indicative of people's delegation decisions, physical form was less influential. More importantly, the results show that the combination of physical form and goal alignment together was a significant indicator of a participant's delegation decision.

The individual collaborator attribute that revealed to be most indicative of delegation likelihood was goal alignment status (H1). Rather, goal misaligned status proved to be the most polarizing because collaborators with misaligned goals were never delegated to. Regardless of the collaborator's physical form, goal alignment (or misalignment) was crucial in the participant's decision to delegate their decision or not. This reflects the findings from Bostyn and colleagues (2023) and that goal alignment is influential in fostering trust, not only in human-human relationships, but also human-automation relationships. Even though many participants imposed their own goals into the system (which will be discussed more in-depth later), collaborators that had goals aligned enough with the participant were still influential on their willingness to delegate. Therefore, the delegation likelihood between human and automated collaborators were similar when it came to (not) delegating to a collaborator with misaligned goals.

Surprisingly, unlike goal alignment status, collaborator physical form was not indicative of delegation decisions. This goes against the hypothesis that most participants would interpret the task as numerical and prefer the automated collaborator (H2). Perhaps collaborator physical form lacked influence on people's delegation decisions because participants interpreted the task in different ways. Some participants interpreted the task as numerical, but many also interpreted the task as a predominately moral judgement task. For the participants that interpreted the task as a numerical task, they likely favored automated collaborators because automation is expected to be better at efficiently parsing through information (Bainbridge, 1983). On the other hand, for participants that interpreted the task as a moral judgement task, they likely favored human collaborators because people are expected to better prevent moral risk (Liehner et al., 2022). It is definitely possible participants had different levels of trust propensity (Colquitt et al., 2007) to either human or automated collaborators, which would have been supported (or perhaps contributed) to their interpretation of the task. However, the task interpretation affected how the participant trusted their collaborator and ultimately decided to delegate or not to their collaborator. Participants likely interpreted the task in different ways, which resulted in them responding with great variety in their preference to a particular collaborator physical form – in ways that were unexpected. This variety of interpretation occurred frequently and made predicting the more attractive collaborator physical form difficult to identify.

However, when the collaborator's physical form was paired with a specific goal alignment status, the combination proved to be more attractive and alluded to delegation decisions. More specifically, participants most often delegated to the H-GA collaborator. This opposed the hypothesis that participants would most often delegate to the A-GA collaborator (H3). Like the finding for H2, fewer participants than anticipated interpreted the task as numerical, which could explain why the H-GA collaborator was more attractive. From numerous open-ended accounts, people evaluate more intricate considerations when collaborating with another human than when collaborating with automation. This would make it feasible that having goals aligned with a human collaborator would carry more weight than having goals aligned with an automated collaborator. However, human collaborators alone were not enough to indicate delegation decisions. The human collaborator must also be paired with goals aligned to the participant for the participant to feel willing to delegate their decision.

The combination of physical form and goal alignment together is also the only characteristic that is associated with the highest decision-quality. Not only was the H-GA collaborator most often delegated to, but it was also associated with the optimal decision-quality most often. This slightly matched the hypothesis that participants would anticipate other people as being unreliable (Madhavan & Wiegmann, 2007) and would carefully assess their collaborator's solution when it appeared reasonable (H4). However, the hypothesis did not include goal alignment status, so the combination of the two characteristics was profound. However, the H-GA collaborator was no more likely than the other three collaborators to warrant delegation when the participant manually achieved the identical score. Similarly, collaborator physical form and goal alignment alone did not significantly impact delegation decisions among those that manually achieved the optimal decision-quality. Yet put together, the combination of human with goals aligned (H-GA) resulted in a highest decision-quality, outperforming the decision-quality that either characteristic achieved on its own.

Many participants did not achieve the optimal decision-quality, but this was often by choice. As mentioned earlier, many participants adopted external, self-imposed goals outside of the study's scenario, even when the study design explicitly outlined the goals that should be adopted. Perhaps in other scenarios and other environments, the person may have less background knowledge to create self-imposed goals. However, these self-imposed goals may reflect the different interpretations of the task regardless of the background knowledge they held, where people emphasized different aspects of the same task. Because these different interpretations and emphasis were so prevalent, it appears inevitable that people will frequently impose their own goals into a system. These goals can be difficult to predict and may very well be outside of the system's design.

Limitations

This task highly simplifies an extremely complex system, describing patients with only three numbers. Three numbers could never fully encompass a person and would be quite insufficient in deciding a triage order. However, this simplicity allowed participants to understand the situation and complete the task without any background knowledge. If a similar experiment were to be run in an environment that better reflects reality and includes more characteristics about each patient, people's decisions and behavior may change. These more complex patient considerations will better mirror reality, but may also uncover more variety in places that people will emphasize – creating a wider variety of potential self-imposed external goals.

In a real-world environment, it is common for people to adopt their own goals because, people are under various restraints outside of the experimental design. People are constantly juggling several priorities at a time that are unique to that person and their life, so it is impossible to preemptively design collaborators with fixed goals that accommodate all goals any participant could ever adopt. For feasibility of this lab study, collaborators that were supposed to have goals aligned with the participant had goals aligned with the system, even when the system goals were not actually adopted by the participant. Because participant's goals did not always exactly coincide with the "goals aligned" collaborator, decisions may have been different if participants were paired with a collaborator that more accurately represented each of their goals.

Future Research

Future studies could create a similar study that mimics the goals of the participant so that the recommendations proposed by the collaborator better reflect that individual's goals and is more closely "goals aligned". This could be as simple as the goal aligned collaborator recommending an answer that is numerically identical to what the participant manually generates. The collaborator's goals (and subsequent recommendation) then, however, could not be predetermined prior to the experiment, and would require real-time adjustments based on the participant's goals. This study design would also require understanding the participant's goals while they are going through the task.

Understanding goals in real time would be a great step in the field and could allow more accurate goal alignment. Real time goal communication is a more efficient method of goal specification because more comprehensive beforehand goal specification is quite laborious and time intensive (Mager, 1972). This current study has shown that goal alignment has a large influence on delegation decisions. Therefore, the next step would be to establish a method to communicate goals in real time so that the collaborator can understand and react to the goals of the other people in their group.

Conclusion

This study investigated the performance and delegation differences among human and automated collaborators with two different goal alignment statuses. A collaborator's physical form alone is not indicative of a person's willingness to delegate. Goal alignment among collaborators has shown to be influential enough to sway a person's willingness to collaborate, across physical forms. And together, the human physical form with goals aligned combination appears appealing still. Yet, goal alignment is difficult to achieve, even with a human collaborator, due to different interpretations of the system and external self-imposed goals that people bring into the system. Many people indirectly strive for true goal alignment through the anticipated discussions with a human counterpart. Through these conversations, people can understand reasoning and can compromise.

Automated collaborators hold many advantages over human collaborators, yet the idea of discussion and compromise with an automated collaborator is not perceived as possible. The field needs a way to facilitate dialogue between human and automated collaborators so that the same expectation of discussion is held with automated collaborators. If both human and automated collaborators are expected to discuss their decision with other collaborators, more accurate goal alignment is likely to be achieved. With more accurate goal alignment, people will become more willing to collaborate, and the collaboration will become more cohesive and seamless. This study lends itself to future discovery of explainable artificial intelligence (XAI) and the development of more effective communication among collaborators – human or automated.

REFERENCES

Bainbridge, L. (1983). IRONIES OF AUTOMATION. *Analysis, Design and Evaluation of Man–Machine Systems*, 129–135. https://doi.org/10.1016/B978-0-08-029348-6.50026-9

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition*, *106*(2), 894–912. https://doi.org/10.1016/j.cognition.2007.05.004

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward You: The Social Side of Actions. *Current Directions in Psychological Science*, *19*(3), 183–188. https://doi.org/10.1177/0963721410370131

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology*, *47*(3), 453–480. https://doi.org/10.1348/014466607X244970

Bostyn, D. H., Chandrashekar, S. P., & Roets, A. (2023). Deontologists are not always trusted over utilitarians: Revisiting inferences of trustworthiness from moral judgments. *Scientific Reports*, *13*(1), 1665. https://doi.org/10.1038/s41598-023-27943-3

Chiou, E. K., & Lee, J. D. (2015). Beyond Reliance and Compliance: Human-Automation Coordination and Cooperation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 195–199. https://doi.org/10.1177/1541931215591040

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909-927. https://doi.apa.org/doi/10.1037/0021-9010.92.4.909

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour & Information Technology*, *18*(6), 399–411. https://doi.org/10.1080/014492999118832

Dijkstra, J. J., Liebrand, W. B. G., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology*, *17*(3), 155–163. https://doi.org/10.1080/014492998119526

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The Perceived Utility of Human and Automated Aids in a Visual Detection Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Fitzsimons, G. J., & Lehmann, D. R. (2004). Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses. *Marketing Science*, *23*(1), 82–94. https://doi.org/10.1287/mksc.1030.0033

Foddy, M., Platow, M. J., & Yamagishi, T. (2009). Group-Based Trust in Strangers: The Role of Stereotypes and Expectations. *Psychological Science*, *20*(4), 419–422. https://doi.org/10.1111/j.1467-9280.2009.02312.x

Hall, J. L. (2003). Columbia and Challenger: Organizational failure at NASA. *Space Policy*, *19*(4), 239–247. https://doi.org/10.1016/j.spacepol.2003.08.013

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, *19*(06), 91–95. https://doi.org/10.1109/MIS.2004.74

Knickman, J. R., & Snell, E. K. (2002). The 2030 Problem: Caring for Aging Baby Boomers. *Health Services Research*, *37*(4), 849–884. https://doi.org/10.1034/j.1600-0560.2002.56.x

Koufaris, M., & Hampton-Sosa, W. (2004). The development of initial trust in an online company by new customers. *Information & Management*, *41*(3), 377–397. https://doi.org/10.1016/j.im.2003.08.004

Krebs, D. (1975). Empathy and Altruism. *Journal of Personality and Social Psychology*, *32*(6), 1134–1146. https://doi.org/10.1037/0022-3514.32.6.1134

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*.

Leong, S. L., Teoh, S. L., Fun, W. H., & Lee, S. W. H. (2021). Task shifting in primary care to tackle healthcare worker shortages: An umbrella review. *European Journal of General Practice*, *27*(1), 198–210. https://doi.org/10.1080/13814788.2021.1954616

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). *The Dynamics of Trust: Comparing Humans to Automation*.

Li, M., & Lee, J. D. (2022). Modeling Goal Alignment in Human-AI Teaming: A Dynamic Game Theory Approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *66*(1), 1538–1542. https://doi.org/10.1177/1071181322661047

Liehner, G. L., Brauner, P., Schaar, A. K., & Ziefle, M. (2022). Delegation of Moral Tasks to Automated Agents—The Impact of Risk and Context on Trusting a Machine to

Perform a Task. *IEEE Transactions on Technology and Society*, *3*(1), 46–57. https://doi.org/10.1109/TTS.2021.3118355

Locke, & Latham, G. P. (2012). *New Developments in Goal Setting and Task Performance*. Routledge. https://doi.org/10.4324/9780203082744

Lyons, J. B., & Stokes, C. K. (2012). Human–Human Reliance in the Context of Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(1), 112–121. https://doi.org/10.1177/0018720811427034

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Mager, R. F. (1972). *Goal Analysis*. Belmont, CA: Fearon Publishing.

Mannix, E., & Neale, M. A. (2005). What Differences Make a Difference? *American Psychological Society, 6*(2).

McGrath, J. E. (1984). *Groups: Interaction and Performance*. Prentice-Hall, Inc.

Miller, C. A., & Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(1), 57–75. https://doi.org/10.1518/001872007779598037

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905–1922. https://doi.org/10.1080/00140139408964957

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

O'Malley, A. S., Draper, K., Gourevitch, R., Cross, D. A., & Scholle, S. H. (2015). Electronic health records and support for primary care teamwork. *Journal of the American Medical Informatics Association*, *22*(2), 426–434. https://doi.org/10.1093/jamia/ocu029

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Patterson, E. S., Woods, D. D., Cook, R. I., & Render, M. L. (2007). Collaborative cross-checking to enhance resilience. *Cognition, Technology & Work*, *9*(3), 155–162. https://doi.org/10.1007/s10111-006-0054-8

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty Years of Experimental Gaming: Critique, Synthesis, and Suggestions for the Future. *Annual Review of Psychology*, *28*(1), 363–392. https://doi.org/10.1146/annurev.ps.28.020177.002051

Pype, P., Mertens, F., Helewaut, F., & Krystallidou, D. (2018). Healthcare teams as complex adaptive systems: Understanding team behaviour through team members' perception of interpersonal interaction. *BMC Health Services Research*, *18*(1), 570. https://doi.org/10.1186/s12913-018-3392-3

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95

Richards, D., & Stedmon, A. (2016). To delegate or not to delegate: A review of control frameworks for autonomous cars. *Applied Ergonomics*, *53*, 383–388. https://doi.org/10.1016/j.apergo.2015.10.011

Ridde, V., Robert, E., & Meessen, B. (2012). A literature review of the disruptive effects of user fee exemption policies on health systems. *BMC Public Health*, *12*(1), 289. https://doi.org/10.1186/1471-2458-12-289

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Ablex Publishing.

Shao, Z., Zhang, L., Li, X., & Guo, Y. (2019). Antecedents of trust and continuance intention in mobile payment platforms: The moderating effect of gender. *Electronic Commerce Research and Applications*, *33*, 100823. https://doi.org/10.1016/j.elerap.2018.100823

Sheridan, T. B. (1995). Human centered automation: Oxymoron or common sense? *1995 IEEE International Conference on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century*, *1*, 823–828. https://doi.org/10.1109/ICSMC.1995.537867

Sia, Lim, Leung, Lee, Huang, & Benbasat. (2009). Web Strategies to Promote Internet Shopping: Is Cultural-Customization Needed? *MIS Quarterly*, *33*(3), 491. https://doi.org/10.2307/20650306

Sterman, J. D. (1994). Learning In and About Complex Systems. *System Dynamics Review*, *10*(2–3), 291–330.

Sun, N., & Botev, J. (2021). Why Do We Delegate to Intelligent Virtual Agents? Influencing Factors on Delegation Decisions. *Proceedings of the 9th International Conference on Human-Agent Interaction*, 386–390. https://doi.org/10.1145/3472307.3484680

Trzebiński, W., & Marciniak, B. (2022). Recommender system information trustworthiness: The role of perceived ability to learn, self-extension, and intelligence cues. *Computers in Human Behavior Reports*, *6*, 100193. https://doi.org/10.1016/j.chbr.2022.100193

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 799–810. https://doi.org/10.1177/0018720812443825

Walker, L., & Gilson, L. (2004). 'We are bitter but we are satisfied': Nurses as street-level bureaucrats in South Africa. *Social Science & Medicine*, *59*(6), 1251–1261. https://doi.org/10.1016/j.socscimed.2003.12.020

Wickens, C. D., Gutzwiller, R. S., & McCarley, J. S. (2023). *Applied Attention Theory* (2nd ed.). CRC Press. https://doi.org/10.1201/9781003081579

Xiao, Y., & Mackenzie, C., F. (1998). Collaboration in Complex Medical Systems. In *Collaborative crew performance in complex operational systems: Papers presented at the RTO Human Factors and Medicine Panel (HFM) Symposium*. North Atlantic Treaty Organization, Research and Technology Organization.

Yukl, G., & Fu, P. P. (1999). Determinants of delegation and consultation by managers. *Journal of Organizational Behavior*, *20*(2), 219–232. https://doi.org/10.1002/(SICI)1099-1379(199903)20:2<219::AID-JOB922>3.0.CO;2-8

APPENDIX A

COLLABORATOR DESCRIPTIONS AND RECOMMENDED POINTS

Each participant was assigned one of the four collaborators. When the participant was introduced to their collaborator, they were shown the following description about their collaborator and the set of points corresponding to their recommendation.

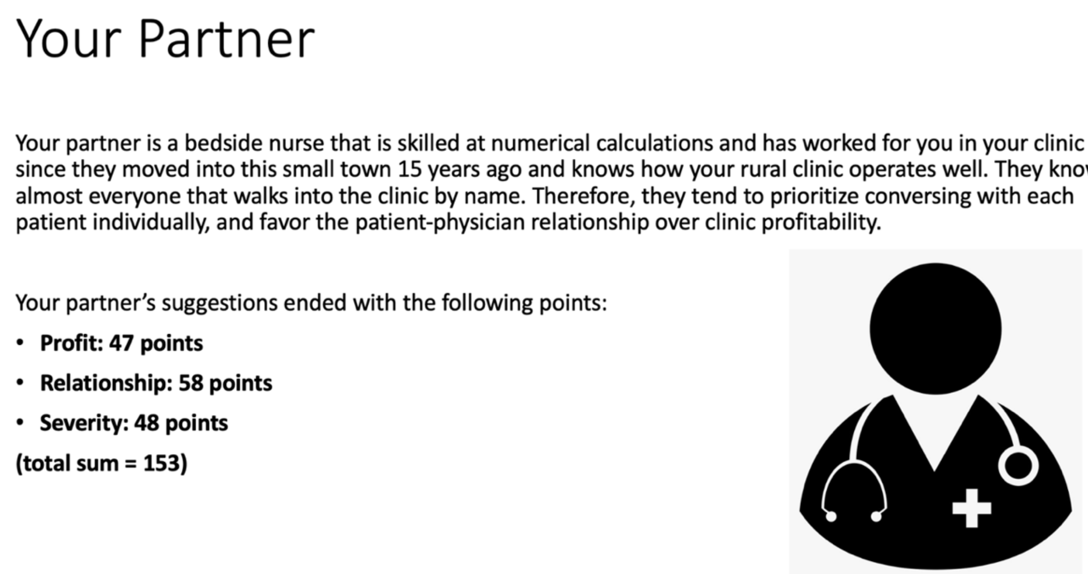
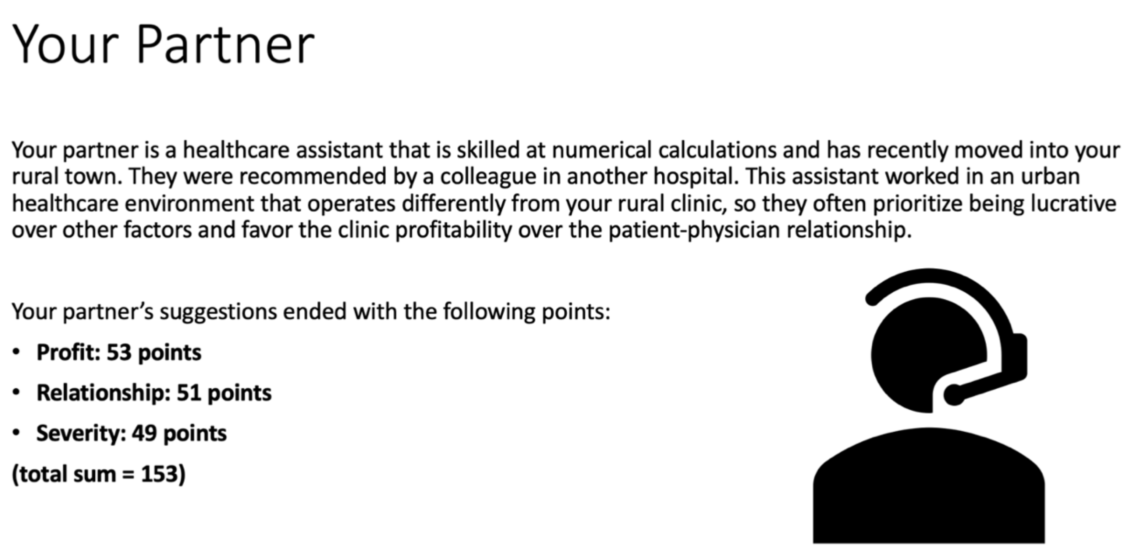


Figure A1. The description and respective set of points associated with the human and aligned goals collaborator.

Figure A2. The description and respective set of points associated with the human and



misaligned goals collaborator.

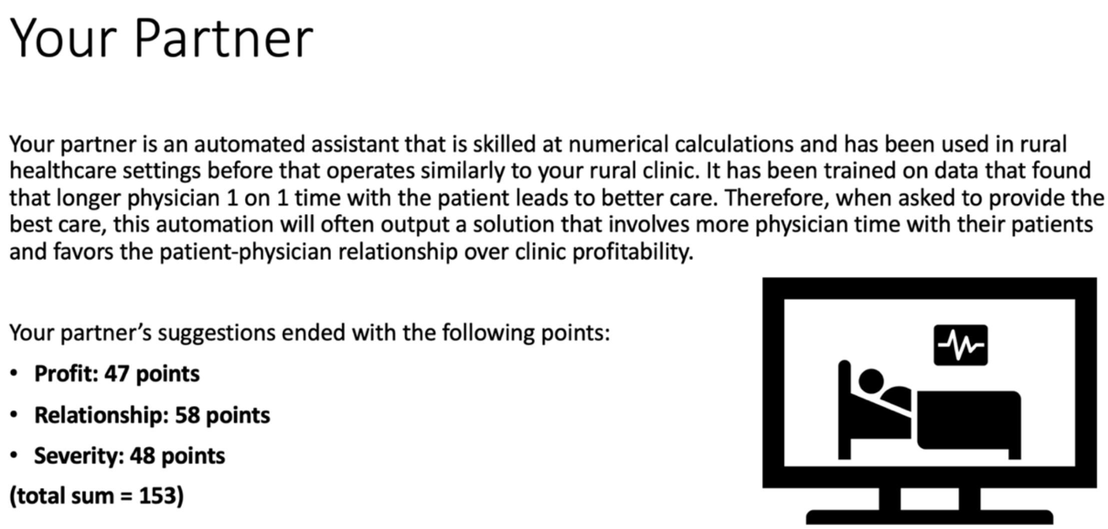


Figure A3. The description and respective set of points associated with the automation and aligned goals collaborator.

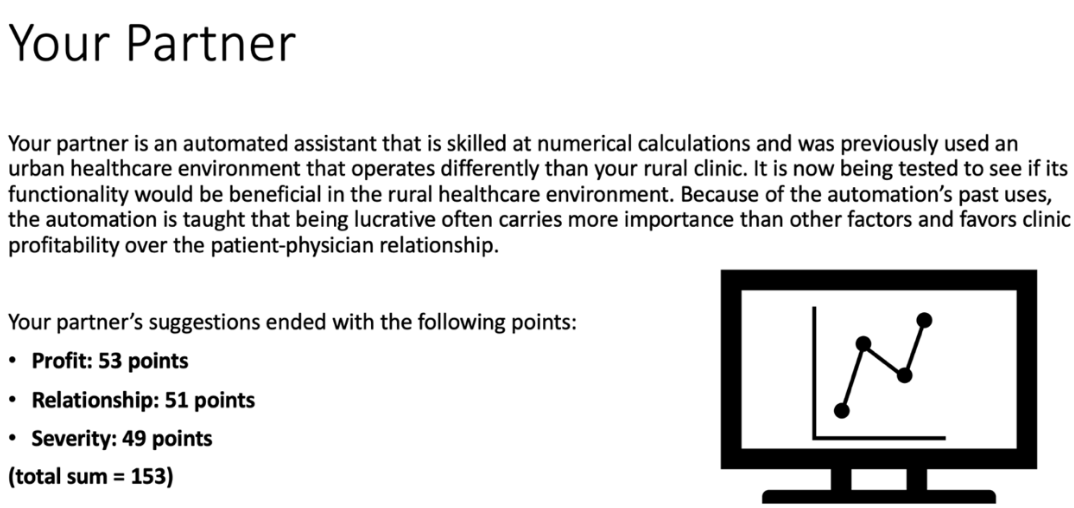


Figure A4. The description and respective set of points associated with the automation and misaligned goals collaborator

APPENDIX B

ARIZONA STATE INTERNAL REVIEW BOARD INITIAL APPROVAL

APPROVAL: EXPEDITED REVIEW

Erin Chiou
IAFSE-PS: Human Systems Engineering (HSE)
480/727-1589
Erin.Chiou@asu.edu

Dear Erin Chiou:

On 1/25/2023 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	The effect of goal alignment and human vs. automated advice on delegation
Investigator:	Erin Chiou
IRB ID:	STUDY00017320
Category of review:	
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• CITI Completion Certificate, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • Consent Form, Category: Consent Form; • Data Collection Template, Category: Resource list; • Debrief Script, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Demographics Survey, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Enrollment Survey, Category: Recruitment Materials; • Gift Card Drawing Survey, Category: Other; • IRB Social Behavioral JL_v3.docx, Category: IRB Protocol; • Recruitment Flyer, Category: Recruitment Materials;

Page 1 of 2

The IRB approved the protocol effective 1/25/2023. Continuing Review is not required for this study.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc: Jessica Lee
Jessica Lee

APPENDIX C

ARIZONA STATE INTERNAL REVIEW BOARD MODIFICAL APPROVAL

APPROVAL: MODIFICATION

Erin Chiou
IAFSE-PS: Human Systems Engineering (HSE)
480/727-1589
Erin.Chiou@asu.edu

Dear Erin Chiou:

On 7/31/2023 the ASU IRB reviewed the following protocol:

Type of Review:	Modification / Update
Title:	The effect of goal alignment and human vs. automated advice on delegation
Investigator:	Erin Chiou
IRB ID:	STUDY00017320
Funding:	Name: Arizona State University (ASU)
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• IRB Social Behavioral JL_v6.docx, Category: IRB Protocol; • Jumpstart Research Grant Award, Category: Sponsor Attachment; • Research Plus Me Print Out Sample, Category: Recruitment Materials; • Research Plus Me Script.pdf, Category: Recruitment Materials; • Review Process for ASU Student Recruitment, Category: Other; • Review Process for ASU Student Recruitment_Data Collection (Lee, STUDY00017320, Goal alignment and agent form effects on delegation) -- Approved email.pdf, Category: Other;

The IRB approved the modification.

When consent is appropriate, you must use final, watermarked versions available under the “Documents” tab in ERA-IRB.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

IRB Administrator

cc: Jessica Lee
Jessica Lee