Perceived Robot Capability

Elizabeth Cha, Anca D. Dragan and Siddhartha S. Srinivasa[1]

Abstract—**Robotics research often focuses on increasing robot capability. If end users do not perceive these increases, however, user acceptance may not improve. In this work, we explore the idea of perceived capability and how it relates to true capability, differentiating between physical and social capabilities. We present a framework that outlines their potential relationships, along with two user studies, on robot speed and speech, exploring these relationships. Our studies identify two possible consequences of the disconnect between the true and perceived capability: (1) under-perception: true improvements in capability may not lead to perceived improvements and (2) over-perception: true improvements in capability may lead to additional perceived improvements that do not actually exist.**

I. INTRODUCTION

Roboticists often focus on increasing robot capability: we make robots faster [1], help them perceive the world [3], and enable them to interact with people [4]. The goal is to make robots a part of everyday life, with purposes ranging from entertainment to assisting with tedious or dangerous tasks [5]–[10].

An integral part of making this goal a reality is *acceptance* [11], [12], which is largely affected by users' *perceptions* [12]. Therefore, increasing robot capability does not necessarily lead to increased acceptance since it is actually the users' perception of capability — the robot's *perceived capability* — that determines acceptance.

For example, imagine a robot designed to clean the home. If a user perceives the robot's capability to be lacking, they may be reluctant to allow it to handle their possessions. On the other hand, if the user overestimates the robot's capability, this can lead to unmet expectations and disappointment.

In this paper, we focus on the idea of perceived capability, and the disconnect between it and the robot's true capability. We start by conjecturing a framework that links *true* and *perceived* capability and then outline their possible relationships. Within this framework, we investigate the effects of two important robot capabilities, *speed* and *speech*, on perceived capability via two user studies.

Framework for Perceived Capability

Prior work in anthropomorphism, mental models, and sensemaking shows there is often a disconnect between users' perceptions and a robot's true capability [13]–[16]. This occurs largely because people are unfamiliar and lack experience with robots. Hence, their knowledge is often based on popular culture which depicts a wide variety of robots with a multitude of capabilities.

In order to investigate this disconnect, we introduce a framework which enumerates the possible relationships between true and perceived capability. Our framework distinguishes between *physical* capability (e.g., doing laundry) and *social* capability (e.g., understanding what someone is saying), and between a particular *skill* and overall capability.
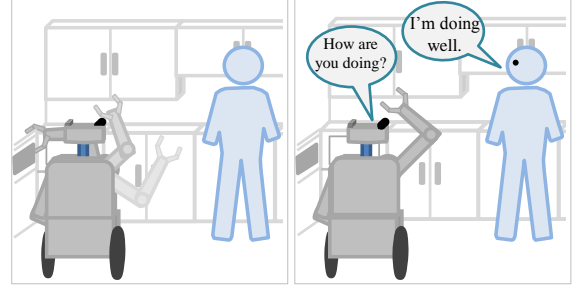
[1]Elizabeth Cha, Anca D. Dragan and Siddhartha S. Srinivasa are with the Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States `lizcha, adragan, siddh@cs.cmu.edu`

Fig. 1: We analyze the effects of two different robot skills — speed and speech — on the robot's perceived capability.

Study on Perceived Capability and Speed

Within this framework, we conduct a user study to investigate the effects of robot speed on perceived capability. Although a considerable amount of robotics research focuses on enabling robots to complete tasks at a reasonable pace, it is unclear how such increases affect our perceptions and hence, attitudes towards robot.

Our findings show that doubling the speed of the robot only slightly increases perceived capability. Comments from users indicate this *under-perception* is because although users perceive the significant increase in the robot's speed, they do not necessarily associate speed with overall capability.

Study on Perceived Capability and Speech

Our second user study focuses on the effects of true skills, such as speaking or succeeding at a physical task, on perceived capability. Speech plays an important role in communication and socialization and is known to cause anthropomorphization, or the process of attributing human-like behaviors and intelligence [17]. Since many robots already utilize speech, we are interested in exploring how its use impacts perceptions of capability.

We investigate two types of speech (Fig.1): *conversational*, or speech which is purely phatic in nature, and *functional*, or speech which is task-oriented and easy generate.

Our findings confirm that a true physical capability difference (succeeding at the task) can affect perceived physical capability and that a true social capability difference (speaking) can affect perceived social capability. However, we also find that the two dimensions can influence each other, leading to *over-perception*: speech, in particular conversational speech, increases perceived *physical* capability, despite adding nothing to the robot's true physical capability.

Overall, we find two possible consequences of a disconnect between true and perceived capability. The first is under-perception: increasing a robot's true capability does not lead to an improvement in users' perceptions, as seen in our study on robot speed. The second is over-perception: increasing a robot's true capability can lead to a much larger improvement in users' perceptions, as seen in our study on speech.

These findings indicate that understanding the effects of true capability can help guide future research by showing what capabilities roboticists should prioritize and how to avoid raising users' expectations to an unrealistic level.
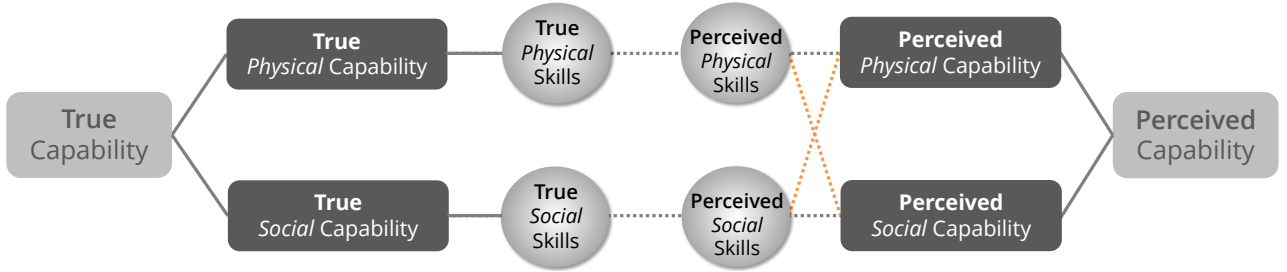
Fig. 2: Framework for perceived capability.

II. PERCEIVED CAPABILITY

To explore the disconnect between true and perceived capability, we conjecture a framework, shown in Fig.2, which looks at capability along three dimensions: true vs. perceived capability, physical vs. social capability, and skill vs. overall capability.

A. True vs. Perceived Capability

True capability refers to what the robot is actually able to do, while *perceived capability* refers to what users believe the robot can do.

Our first user study (Section III) suggests improvements in true capability do not always translate to increases in perceived capability. Furthermore, our next study (Section IV) reveals that improvements in true capability can lead to a disproportionate increase in perceived capability. Therefore, it important to understand how to convey a robot's true capability in order to set the correct expectations.

Robots are a novel and diverse technology which are not yet widely available. Thus, preconceptions of their capabilities are unlikely to be realistic as most people's knowledge of robots come from science fiction and popular culture. When encountering a real robot, people may try to reason about it by comparing it to a similar entity, such as a human [18]. A growing body of research suggests this is especially true for robots with more human-like appearances or behaviors [14], but this can be problematic as robots and humans have vastly different strengths and weaknesses [19].

Moreover, although we often compare robots and humans, prior work shows we still treat them quite differently [20], [21]. Therefore, it is important to understand how people perceive a robot's skills and effects of these perceptions.

B. Physical vs. Social Capability

Most tasks desired from future robots can be divided into two distinct categories: *physical* and *social*.

1) Physical: Perceived physical capability is intimately related to trust. When the robot makes a mistake, trust drops [22] because the robot does not meet expectations of reliability which suggests that its perceived physical capability is too high. In our study in Section IV, we also use physical mistakes to manipulate true capability.

Drawing from the large body of survey literature, we developed a method for evaluating perceived physical capability [7], [23]–[25]. We identified fourteen common tasks (see Table I) that people envision future robots to perform, with the majority of tasks taking place in personal settings such as the home. This corresponds with [5], who found people were most open to robots acting as housekeepers or assistants.

To measure perceived physical capability, we ask users to rate how capable the robot is of performing each task on a 7-point Likert scale.

2) Social: Socially interactive robots are valued for their potential as companions or informational assistants [26]. However, true social robots require human-like intelligence, a longstanding goal in the field of AI [17], [27]. Because this is still an open research problem, several people have adopted the weak AI stance, concentrating instead on projecting the illusion of intelligence [17], [28].

Projecting intelligence simply requires displaying the attributes that facilitate people's definition of intelligence [17]. Even the Turing Test which examines a machine's ability to exhibit behavior indistinguishable from a human, is based on perceived rather than true intelligence. Much of what is considered to constitute human intelligence actually relates to social ability: thus, we also explore the notion of *perceived social capability*.

[17] argues that perceived social capability may be a more valuable tool for acceptance then true social capability. However, if we manipulate perceived capability too far in either direction, we may hinder acceptance. On the one hand, if expectations are set too high, the robot may end up disappointing users. On the other hand, if expectations are too low, the robot may fall into disuse. Thus, it is important to strike just the right balance.

Nomura's Robot Anxiety Scale (RAS), one of the most popular robot scales, explores difficulties that can arise from interacting with a communicative robot [29]. Many items in this scale concentrate on not just communication, but socialization. Drawing inspiration from the RAS, we developed eleven items that focus on a robot's ability to successfully communicate and interact with humans, shown in Table I.

To measure perceived social capability, we ask users to rate how capable the robot is of performing each task on a 7-point Likert scale.

TABLE I: Perceived Capability Measures

Physical	Social
wiping down cleaning surfaces	giving information and news
cleaning up the dining table	understanding what I am talking about
taking out the trash	understanding my emotions
tidying up	informing me when something is wrong
organizing the pantry	following the direction of a conv.
doing laundry	understanding difficult conversations
preparing simple meals	informing me what it is about to do
doing lawn work	making me understand what it saying
bringing a drink	understanding what I am doing
helping me get ready	talking about irrelevant things*
acting as alarm system	-
watering/caring for plant	-
pet care	-
child care	-

Fig. 3: Videos shown to participants depicted the robot retrieving a microwave meal with an actor washing dishes nearby. The robot 1) comes into the kitchen, 2) opens the microwave, 3), retrieves the microwave meal, 4), closes the microwave and 5) leaves the kitchen.

C. Skill vs. Overall Capability

Capability encompasses the entirety of what a robot can do, making it difficult to directly measure and manipulate. Thus, we deconstruct capability into individual *skills* or abilities that a robot can possess such as folding clothes or telling jokes. Skills, like capability, can be either physical or social, and can be true or perceived.

D. Relationships

Our framework explores true and perceived capability via a sequence of connections, shown in Fig.2. Some connections are set, while others, shown by dotted lines, are not always present, or are weaker than expected.

A robot's true capability is comprised of its true physical and social capability which are each comprised of true skills. However, moving forward through the framework, improvements in true skills might not necessarily be reflected in perceived skills, due to observability. For example, a robot gains new sensors to record audio, but a user observing the robot may not notice the difference. Similarly, improvements in perceived skills might not lead to improvements in perceived capabilities, due to users' definitions off capability.

While some connections may be weaker than expected, other can be stronger. One of the contributions of our study (in Section IV) is to show the existence of such a connection, depicted in orange in Fig.2. This can be attributed to the complex processes that occur when we interact with robots, such as anthropomorphization, which can cause users to attribute non-existent abilities to the robot, e.g. conversational speech can lead to the perception of capability at physical tasks.

Thus, there are potential connections between perceived social skills and perceived physical capability, or between perceived physical skills and perceived social capability. These connections can be direct or can be caused by connections between a perceived skill (like speech) and another perceived skill (like understanding natural language).

E. Measuring Perceived Capability

We measure perceived physical and social capability through the tasks in Table 1. For each task, we present users with the following statement: "I believe the robot is capable of X," where X is the task. Users are asked to rate their agreement with the statement on a 7-point Likert scale. Users also asked when comparing two robots to choose the robot they believe is more capable of performing each task.

III. SPEED STUDY

Our first study is about under-perception. An increase in true capability does not necessarily lead to an increase in perceived capability.

We exposed users to two videos of a robot performing a physical task next to an actor. In each scenario, the robot enters the kitchen and retrieves a microwave meal while standing next to the actor. In one video, the robot moves faster and spends less time planning.

A. Study Design

1) Manipulated Variables: We manipulated speed by producing two videos of a robot performing a physical task — retrieving a microwave meal — at two different speeds (2.3 minutes vs. 1.15 minutes to complete the task). The lower speed represents the state of the art, while the higher speed represents a capability well beyond current robot manipulation skills in both planning and execution time.

2) Participants: We recruited 20 participants (12 females and 8 males) through Amazon Mechanical Turk. All participants were located in the United States, primary English speakers and ranging in age from 23 to 65 (M=39.0, SD=10.55) years. 40% percent of participants were male and 60% were female. Participants were compensated $4 for successful completion of the study.

Participants were told that they were taking part in a survey to design better home robots. All participants successfully answered a set of control questions about the videos they were shown and none had previously participated in a study with the robot.

This resulted in 14 participants spread across the 2 conditions. On average, participants rated their familiarity with robots as 2.2 (SD=1.28) and their previous level of interaction as 1.3 (SD=0.47) on a 7-point Likert scale.

3) Procedure: We opted for a *within-subjects design*, where participants were shown both the slow and fast videos, in order to enable direct comparisons and stay away from absolute ratings (as suggested by a pilot study). The order of the videos shown were counterbalanced to negate ordering effects.

Participants were given a link through Amazon Mechanical Turk to the study. After reading the instructions and giving their consent, participants were shown the first video of the robots labeled "Robot 1." To continue the study, participants had to watch the entire video. They then answered questions about the video and perceived capability. They repeated this process after watching the second video.

4) Hypothesis: We hypothesize that even when doubling the speed of the robot, it will not affect perceived capability as the robot is still much slower than a human.

5) Measures: We used the perceived capability measures outlined in Section II-E.

B. Results

Our hypothesis predicted that speed would not significantly affect a robot's perceived physical capability. To

test this hypothesis, we performed a non-inferiority test to show that the difference between the slow and fast robot is significantly greater than a negative margin $-\Delta$. Setting $\Delta = 5\%$, a one-tailed paired t-test supported the hypothesis: $-0.296 > -0.3$, $t(25) = -1.796, p = 0.042$.
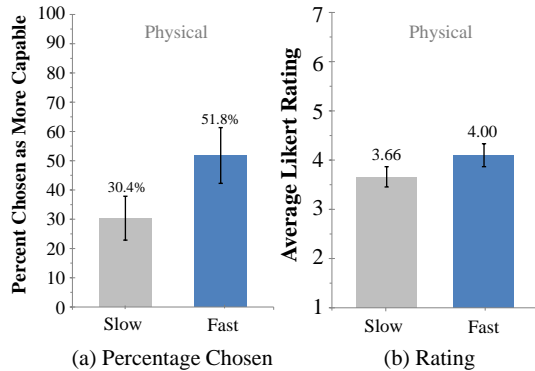

Fig. 4: Perceived physical capability metrics for slow vs. fast robot.

Despite a difference in true capability that would be considered very large from a robotic algorithmic standpoint, the mean ratings were not significantly different (Fig.4). Since speed is a physical trait, we did not expect it to affect perceived social capability: we found no significant difference in the perceived social capability ratings.

We confirmed in a follow-up question that users did perceive the difference in speed, but this had a very small effect on the robot's perceived capability, i.e. on their perceptions of what tasks the robot is capable of performing. They said, for example, that despite the robot being faster, it was either "still not fast enough", or it did not matter, because the "robot's speed did not impact its ability to do other tasks" and thus, cannot be trusted.

These results suggest that true improvements in capability do not necessarily lead to perceived improvements in capability: there can be a disconnect between the two.

IV. SPEECH STUDY

Speech is an interesting and uniquely human behavior. We choose to investigate it within our framework for perceived capability because of its continual usage and popularity in robotics.

A. Background

Among many possible skills that can affect perceived capability, we focus on *speech* because it causes anthropomorphization [14] and plays an important role in human communication and socialization. However, anthropomorphization may cause users to perceive capabilities that do not actually exist: this is a main finding of our study in the next section.

Starting from early chatterbots such as ELIZA and PARRY, much of speech research has focused on enabling machines to make *conversation* with humans [30], [31]. However, making conversation is difficult, because it relies on grounding — it requires participants to coordinate both the content and process of the communication [32]–[34]. As a result, a different type of speech is predominant in current technology: a more rudimentary, *functional* speech that arises

from people's preference for interacting via speech [5]. This type of speech is easy to generate and consists mostly of relaying a machine's status or responding to a set of pre-programmed voice commands.

Motivated by the different types of speech in research and the real world, our study divides speech into these two categories: *functional* and *conversational*.

Functional Speech. We define functional speech as speech that relates purely to the state of the robot. This speech utilizes warnings and requires no human-understanding, making it easy to generate.

Traditional industrial robots relay information with methods such as blinking lights or warning sounds [35]. Although these are effective, speech allows a larger range of information to be communicated and can be helpful in appropriately setting users' expectations [36].

Fig.5 shows an example of functional speech.

Conversational Speech. We define conversational speech as purely phatic or social in its purpose. Unlike functional speech, it requires human-understanding, making the robot responsive and engaging.

Although true conversational speech often conveys information about the speaker, we simplify it in this study to be purely phatic, or social in its purpose, containing no information about the robot's intended task [37], [38].

Fig.5 also shows an example of conversational speech.

B. Study Design

In order to further investigate the relationship between true and perceived capability, we conducted an online user study looking at the effects of success in completing a physical task and two types of speech: functional and conversational.

We exposed users to different scenarios of the robot performing a physical task next to an actor shown through video. In each scenario, the robot enters the kitchen and retrieves a microwave meal while standing next to the actor. In half of the scenarios, the robot fails to complete the task by dropping the microwave meal on the ground. The robot then leaves the kitchen, possibly with the meal.

In the scenarios, the robot also varies its use of speech. In some cases, the robot engages the actor in dialogue. The actor only speaks to the robot, if it engages him first.

1) Design Choice: We chose to perform an online study with videos rather than an in person study due to the challenges in controlling all variables of the interaction. Showing the robot performing a difficult task was necessary for one of our manipulated variables, but the robot lacked robustness when performing the task. In several scenarios, the robot and actor engage in controlled dialogue. However, in an in person study, the participants responses to the robot varies the remaining interaction, creating a different experience for each user.

2) Manipulated Variables: We manipulated the presence of **functional speech**, the presence of **conversational speech**, and the robot's **success** in completing a physical task (success vs. failure) for a total of $2 \times 2 \times 2 = 8$ conditions (depicted in Fig.5).

For each condition, we created a script that outlined the robot's actions performing the meal retrieval task as well as the robot and actor's speech. We then created video recordings of the robot and actor following each script.

In the *success* videos, the robot successfully retrieves the meal and leaves the kitchen with it.
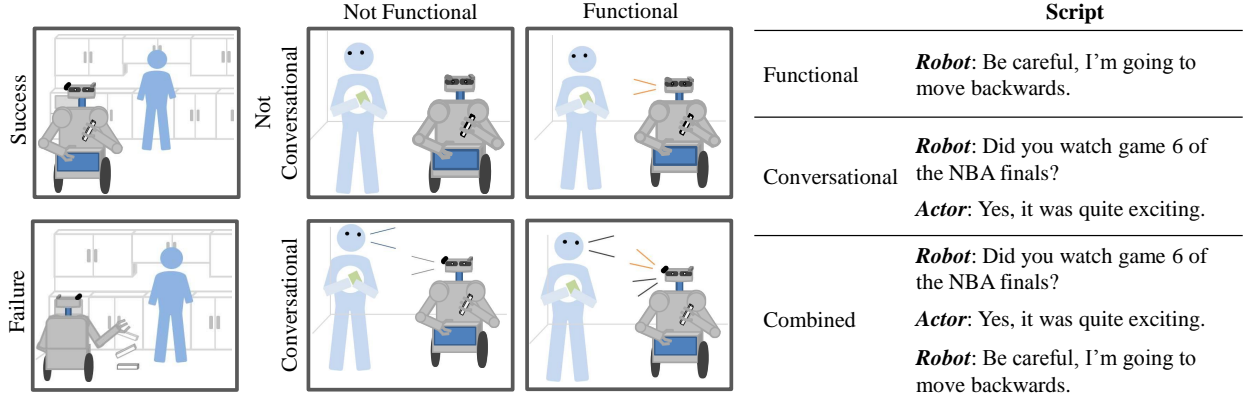
Fig. 5: A visual depiction of the three factors we manipulated (functional speech, conversational speech, and task success), each with two levels.

In the *failure* videos, the robot drops the microwave meal before leaving the kitchen. The robot does not acknowledge dropping the meal in order to maintain consistency.

In the *functional* videos, only the robot speaks, warning the nearby actor of its future movements.

In the *conversational* videos, the robot exchanges dialogue with the actor about an upcoming sports game. The robot does not relay any information about its state other than responding "well" to a query about how it is doing.

In the *combined* functional and conversational videos, the robot uses both functional and conversational speech by combining its sports dialogue with warnings related to its actions.

In the *no speech* videos, the robot does not speak or acknowledge the actor at all.

3) Participants: We recruited 225 participants (98 females and 127 males) through Amazon Mechanical Turk. All participants were located in the United States, primary English speakers and ranging in age from 18 to 61 (M=32.13, SD=9.77) years. 56% percent of participants were male and 44% were female. Participants were compensated $4 for successful completion of the study.

Participants were told that they were taking part in a survey to improve interactions with robots. We eliminated participants who failed to answer a set of control questions about the videos they were shown or who had previously participated in a study with the robot.

This resulted in 186 participants spread across the 28 conditions. Due to random assignment and elimination of participants, there were between six and eight participants in each condition.

On average, participants rated their familiarity with robots as 2.54 (SD=1.55) and their previous level of interaction as 1.66 (SD=1.04) on a 7-point Likert scale.

4) Procedure: We opted for a *mixed study design*, where participants were shown two of the eight videos, in order to enable direct comparisons and stay away from absolute ratings (as suggested by our pilot study).

We took all possible pairwise combinations of the eight videos to get a total of 28 study conditions. Participants were randomly assigned to one of the 28 conditions. In order to negate ordering effects, we randomized the order videos were shown in.

Participants were given a link through Amazon Mechanical Turk to the study. After reading the instructions and giving their consent, participants were shown the two videos of the robots, labeled "Robot 1" and "Robot 2." To progress in the study, participants had to watch the videos from start to finish. After watching the videos, participants answered our questionnaires for perceived physical and social capability. Afterwards, participants gave some general feedback about the robot and themselves.

5) Measures: We used the perceived capability measures outlined in Section II-E.

Hypothesis 1: *When the robot fails to perform a physical task, perceived physical capability decreases.*

Perceived physical capability is measured by participants' beliefs that that a robot can perform these common tasks. We predict that showing the robot fail at one of these tasks will negatively impact perceived physical capability.

Hypothesis 2: *When the robot uses speech, perceived social capability increases.*

Speech plays an important role in communication for humans. The ability to clearly express yourself is key to successful socialization, making speech an import medium for social behavior. We predict that showing the robot speaking will positively impact perceived social capability.

Hypothesis 3: *When the robot uses speech, perceived physical capability increases*

Although speech is not a physical trait, it is known to cause anthropomorphization, which can lead to attributing other human-like traits to the robot [17], [18]. Since the capabilities of robots are relatively unknown to most, some may assume that robots possess physical abilities on par with humans. Thus, we predict that showing the robot speaking will positively impact perceived physical capability.

Hypothesis 4: *Conversational speech increases perceived physical capability more than functional speech.*

True conversational speech requires human intelligence, whereas functional speech is easy to generate and requires no human understanding. People may, therefore, compare a robot that engages in conversational speech to physically capable humans. We thus predict that conversational speech will positively impact perceived physical capability more than functional speech.

C. Results

Fig.6 plots the overall results for the forced choice measure. In the following, we analyze each hypothesis individually.
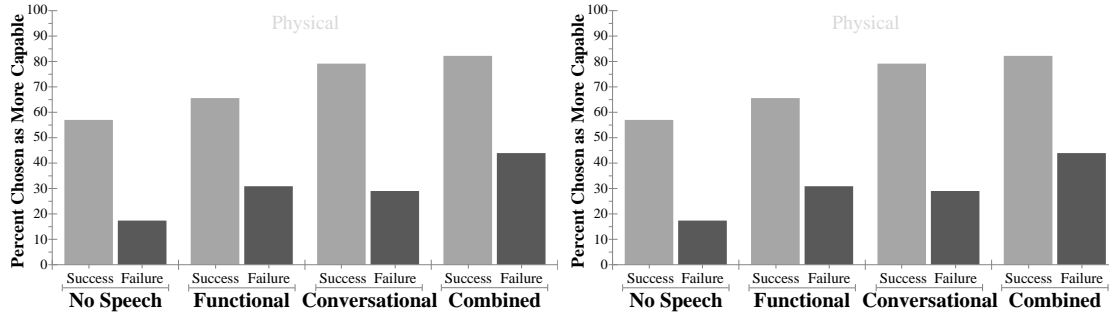
Fig. 6: The percentage of time users chose each of the eight robot types, for both perceived physical capability (left) and perceived social capability (right).

1) Effects of Speech and Success: Hypotheses 1 through 3 predicted the effects of speech (aggregated over conversational and functional) and success on perceived capability. To test these hypotheses, we performed a factorial repeated-measures analysis of variance (ANOVA) for each of our measures: perceived physical capability and perceived social capability.
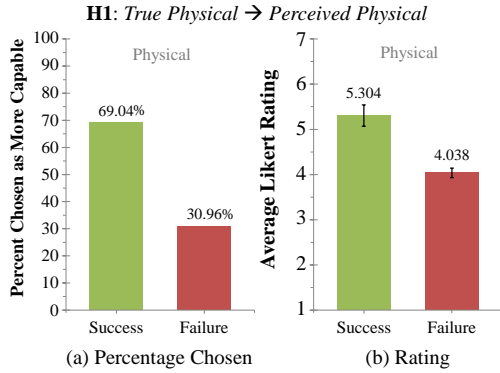

Fig. 7: Perceived physical capability metrics for robots that succeeded vs. robots that failed.

With respect to H1, participants rated robots that were successful 1.267 more than robots that failed to complete the task on the perceived physical capability measure ($F(1, 378) = 97.27$, $p < .001$). They also chose robots that were successful 69% of the time. Fig.7 shows a comparison between the robots that succeed and those that fail.

We did not find a significant difference in the perceived social capability ratings for robots that succeeded compared to robots that failed: *failing at a physical task does not necessarily make participants believe the robot is less socially capable.*

With respect to H2, participants rated robots that used speech, on average, 2.907 more than robots that did not use speech on our perceived social capability measure ($F(1, 378) = 238.03$, $p < 0.001$). Robots that used speech were also chosen as more socially capable 85% of the time. Fig.8 shows a comparison between robots that spoke and those that did not.

Our analysis also supported H3: participants rated robots that were silent 0.387 less in physical capability than robots that spoke ($F(1, 378) = 15.88$, $p < 0.001$). They also chose robots that spoke as more physically capable 60% of the time. Fig.9 shows a comparison between the robots that spoke and those that did not.
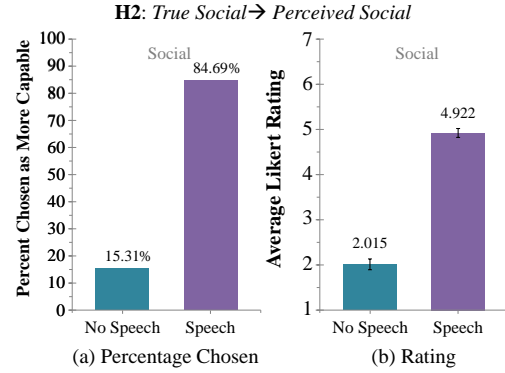

Fig. 8: Perceived social capability metrics for robots that spoke vs. robots that did not.
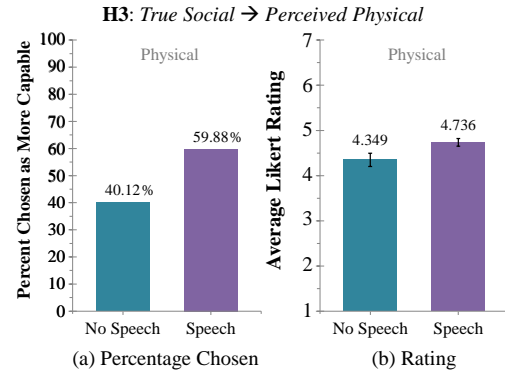

Fig. 9: Perceived physical capability metrics for robots that spoke vs. robots that did not.

2) Different Types of Speech: H4 predicted that conversational speech increases perceived physical capability more than functional speech. To analyze how these two different types of speech affect perceived physical capability, we ran a factorial repeated-measures ANOVA using success, functional speech and conversational speech as factors.

We found significant main effects for all three factors: sucess $F(1, 378) = 129.7$, $p < 0.001$; functional $F(1, 378) = 7.14$, $p < 0.001$; conversational $F(1, 378) = 14.2$, $p < 0.001$.

On average, both types of speech increased perceived physical capability and there was no significant interaction effect. However, when analyzing the mean ratings and direct comparison percentages, we came across a surprising finding:
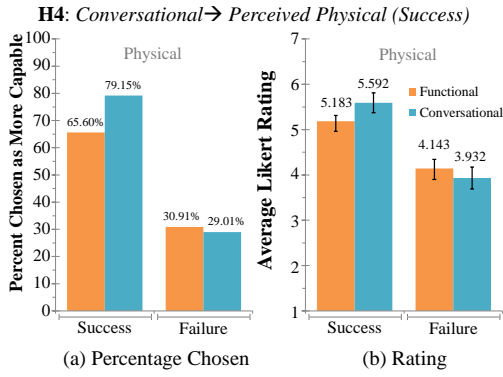
Fig. 10: Perceived physical capability of conversational and functional speech when the robot succeeds and when it fails.

Although conversational speech tends to increase perceived capability more than functional speech when the robot succeeds, this is no longer the case when the robot fails at its task.

In fact, as Fig.10 shows, both the percentage chosen and the mean rating are lower for conversational robots when they fail at the task.

V. DISCUSSION

In this work, we presented a framework for perceived capability and explored within this framework, the effects of speech and speed on perceived capability.

We found that increases in robot speed do not significantly affect perceived physical capability. Although users perceive the increased speed, we discovered they did not associate speed with a robot's physical capability. This could be because humans naturally assume that any reasonable robot will complete tasks at the same pace as a person.

Our results also showed that robot speech positively affects perceived social capability. Interestingly, we discovered that speech also positively affects perceived physical capability, despite true physical capability remaining unchanged.

This could be due to people believing that physical capability is trivial: many of the tasks that make up our physical capability measure are easily completed by humans. Speech, on the other hand, can give the illusion of human intelligence, which may seem more difficult to achieve.

We also looked at two types of speech, functional and conversational, and found that conversational speech increases perceived physical capability *more* than functional speech, but *only when the robot succeeds*. Since conversational speech more closely resembles everyday, human dialogue, users might be more likely to attribute higher levels of intelligence than with simple, functional speech.

Another explanation is that functional speech, through its warnings, reveals the true capability of the robot. Participants commented that the robot's warnings "revealed its limitations" and "made it apparent what the robot could and could not do."

However, when the robot fails, conversational speech has lower perceived physical capability than functional speech. Comments from participants suggest that the robot's ability to make social chit-chat challenges their perceptions when the robot fails. This indicates that users have higher expectations for conversational robots which are challenged when

the robot fails. Participants also pointed out that the robot's inability to acknowledge its error or recover seemed at odds with its relatively human-like speech.

On the other hand, if the robot gave warnings, participants were less surprised when the robot dropped the meal suggesting that functional speech is successful in setting more realistic expectations, closer to the robot's true capability. This was also the case if the robot used both functional and conversational speech. This interpretation follows [36], which found that when robots utilize warnings or mitigation strategies, such as apologizing, people are more forgiving if a breakdown occurs.

Perceived Capability Findings. Looking back to our framework, our findings confirmed that success, a true physical skill, was properly perceived, and as a result affected the perceived physical capability. We also found that a true social skill, speech, affects perceived social capability.

However, speed, a true physical skill was properly perceived but did not affect perceived physical capability.

We were also able to show that there is a cross effect between the physical and social realms, which is part of over-perception — perceiving capability that is not actually there. Speech not only affects perceived social capability, but perceived physical capability as well. We see this cross connection in our framework: a true social skill i ncreases a perceived social skill, which increases perceived social capability *and* perceived physical capability.

One explanation for this connection is that some perceived skills have a direct connection to other perceived skills in the framework. For instance, perceived speech could be affecting a physical perceived skill, and, through it, overall perceived physical capability. This is an area of future exploration: testing this connection by asking users specifically about other perceived skills, as opposed to measuring the overall perceived physical capability.

On the other hand, under-perception can also happen: we saw in our user study on speed (Section III) that it is possible for a skill, such as speed, to not affect either perceived physical or perceived social capability.

Limitations. Speech is currently utilized on many robots, but generating realistic dialogue remains a challenge. This motivated our decision to look at functional speech, which is simplistic, easy to generate and most importantly, resembles the speech used by real world by technology, like cell phones. This created an interesting contrast and allowed us to test a more feasible form of speech.

We also simplified conversational speech to be purely phatic, although in real life, speech is often both informative and social. This helped to maximize the effectiveness of each type of speech as was verified by an initial pilot study.

Furthermore, we used videos in the study and not real interactions. We made this decision to ensure that the robot's fluctuating interaction with a naive user would not introduce confounds into the results.

This work focuses on first impression because they play an important role in acceptance. However, perceptions are often based on our prior knowledge. It would be interesting in the future, to look at how users' experiences and familiarization with a robot alters their perceived capability.

Implications. The disconnect between true and perceived capability can lead to both under-, as well as over-perception.

With under-perception, an increase in true capability will not necessarily be reflected in acceptance, like we saw with

speed. Our results suggest that even if we make robots 5 times faster, there will still be important challenges to acceptance. Studying under-acceptance in more depth could inspire prioritizing robotics research, focusing on the critical factors that will improve acceptance.

With over-perception, an increase in true capability can lead to false perceptions of capability, like we saw with speech. This can be dangerous, because it can set unrealistic expectations: when expectations are not met, trust in the robot drops, which also impacts acceptance [22], [39], [40].

Therefore, mitigating expectations when adding to capability is also important. For instance, our study suggests that functional speech is better at setting perceived capability to a more realistic level. On the other hand, several participants noted how tiresome warnings are.

Overall, our work is a first step in understanding the relationship between true and perceived capability. Although this work presents a simple framework, our findings create the foundation for building a probabilistic graphical model for perceived capability. We are excited to investigate the connections in this graphical model further, and contribute to a better understanding of under- and over- perceptions.

REFERENCES

[1] J. Canny, *The complexity of robot motion planning*. The MIT press, 1988.
[2] M. Dogar and y. S. Srinivasa, journal=Robotics: Science and Systems VII, "A framework for push-grasping in clutter."
[3] G. N. DeSouza and A. Kak, "Vision for mobile robot navigation: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 2, pp. 237–267, 2002.
[4] R. Deits, S. Tellex, P. Thaker, D. Simeonov, T. Kollar, and N. Roy, "Clarifying commands with information-theoretic human-robot dialog," *Journal of Human-Robot Interaction*, 2013.
[5] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, and I. Werry, "What is a robot companion-friend, assistant or butler?" in *IEEE/RSJ IROS*, 2005.
[6] B. News. (2004, Oct.) Robots set to get homely by 2007. [Online]. Available: http://news.bbc.co.uk/2/hi/technology/3764142.stm
[7] C. Ray, F. Mondada, and R. Siegwart, "What do people expect from robots?" in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*. IEEE, 2008, pp. 3816–3821.
[8] C. Stanger, C. Anglin, W. Harwin, and D. Romilly, "Devices for assisting manipulation: a summary of user task priorities," *Rehabilitation Engineering, IEEE Transactions on*, vol. 2, no. 4, pp. 256–265, 1994.
[9] J. Swartz. (2012) Robots are marching into homes. [Online]. Available: http://www.usatoday.com/story/tech/2012/11/12/robots-irobot-romotive-bossa-nova/1695383/
[10] J. E. Young, R. Hawkins, E. Sharlin, and T. Igarashi, "Toward acceptable domestic robots: Applying insights from social psychology," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 95–108, 2009.
[11] J. M. Beer, A. Prakash, T. L. Mitzner, and W. A. Rogers, "Understanding robot acceptance," School of Psychology, Human Factors and Aging Laboratory, Georgia Institute of Technology, Tech. Rep. HFA-TR-1103, 2011.
[12] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS quarterly*, pp. 319–340, 1989.
[13] J. Forlizzi, "How robotic products become social products: an ethnographic study of cleaning in the home," in *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. ACM, 2007, pp. 129–136.
[14] S. R. Fussell, S. Kiesler, L. D. Setlock, and V. Yew, "How people anthropomorphize robots," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. ACM, 2008, pp. 145–152.
[15] A. Powers and S. Kiesler, "The advisor robot: tracing people's mental model from a robot's physical attributes," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 2006, pp. 218–225.
[16] K. E. Weick, *Sensemaking in organizations*. Sage, 1995, vol. 3.
[17] B. R. Duffy, "Anthropomorphism and the social robot," *Robotics and autonomous systems*, vol. 42, no. 3, pp. 177–190, 2003.
[18] S. Kiesler and J. Goetz, "Mental models of robotic assistants," in *CHI'02 extended abstracts on Human Factors in Computing Systems*. ACM, 2002, pp. 576–577.
[19] M. K. Lee, K. P. Tang, J. Forlizz, and S. Kiesler, "Understanding users: Perception of privacy in human-robot interaction," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*. IEEE, 2011, pp. 181–182.
[20] C. Bartneck, C. Rosalia, R. Menges, and I. Deckers, "Robot abuse–a limitation of the media equation," in *Proceedings of the Interact 2005 Workshop on Agent Abuse, Rome*, 2005.
[21] M. K. Lee and M. Makatchev, "How do people talk with a robot?: an analysis of human-robot dialogues in the real world," in *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2009, pp. 3769–3774.
[22] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, and H. Yanco, "Effects of changing reliability on trust of robot systems," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*. IEEE, 2012, pp. 73–80.
[23] S. N. Copleston and G. Bugmann, "Personal robot user expectations," *Robotics and Intell. Syst. Technical Report*, pp. 08–01, 2008.
[24] N. Ezer, A. D. Fisk, and W. A. Rogers, "More than a servant: Self-reported willingness of younger and older adults to having a robot perform interactive and critical tasks in the home," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 53, no. 2. SAGE Publications, 2009, pp. 136–140.
[25] C. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "Older adults preferences for and acceptance of robot assistance for everyday living tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1. SAGE Publications, 2012, pp. 153–157.
[26] S. Andrist, E. Spannan, and B. Mutlu, "Rhetorical robots: making robots more effective speakers using linguistic cues of expertise," in *ACM/IEEE HRI*. IEEE Press, 2013, pp. 341–348.
[27] B. J. Copeland, *Artificial intelligence*. Blackwell Oxford, 1993.
[28] P. Hayes and K. Ford, "Turing test considered harmful," in *IJCAI (1)*, 1995, pp. 972–977.
[29] T. Nomura, T. Kanda, T. Suzuki, and K. Kato, "Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots," *Robotics, IEEE Transactions on*, vol. 24, no. 2, pp. 442–451, 2008.
[30] R. C. Parkinson, K. M. Colby, and W. S. Faught, "Conversational language comprehension using integrated pattern-matching and parsing," *Artificial Intelligence*, vol. 9, no. 2, pp. 111–134, 1977.
[31] J. Weizenbaum, "Elizaa computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.
[32] H. H. Clark and S. E. Brennan, "Grounding in communication," *Perspectives on socially shared cognition*, vol. 13, no. 1991, pp. 127–149, 1991.
[33] P. J. Hinds, T. L. Roberts, and H. Jones, "Whose job is it anyway? a study of human-robot interaction in a collaborative task," *Human-Computer Interaction*, vol. 19, no. 1, pp. 151–181, 2004.
[34] S. Kiesler, "Fostering common ground in human-robot interaction," in *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, 2005, pp. 729–734.
[35] K. Funakoshi, K. Kobayashi, M. Nakano, S. Yamada, Y. Kitamura, and H. Tsujino, "Smoothing human-robot speech interactions by using a blinking-light as subtle expression," in *Proceedings of the 10th international conference on Multimodal interfaces*. ACM, 2008, pp. 293–296.
[36] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski, "Gracefully mitigating breakdowns in robotic services," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 203–210.
[37] B. Malinowski, "The problem of meaning in primitive languages," *Language and literacy in social practice: A reader*, pp. 1–10, 1994.
[38] V. Zegarac, "What is" phatic communication"?" *Pragmatics AND Beyond New Series*, pp. 327–362, 1998.
[39] M. Desai, K. Stubbs, A. Steinfeld, and H. Yanco, "Creating trustworthy robots: Lessons and inspirations from automated systems," *Proc. Artificial Intelligence and Simulation of Behaviour Conv.: New Frontiers in Human-Robot Interaction, Scotland*, 2009.
[40] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*. IEEE Press, 2013, pp. 251–258.