Knowing End User Affects Design Process: Argument for Robot Factors Design

A thesis submitted by

Aleksandra Kaszowska

in partial fulfillment of the requirements for the degree of

Master of Science
in
Psychology

Tufts University
May 2017

Advisor: Holly A. Taylor, Ph.D.
Examining Committee: Ariel Goldberg, Ph.D., William C. Messner, Ph.D.

Abstract

Modern robotic systems can be reprogrammed to complete a wide range of tasks, requiring engineers to approach robots as both tools and tool operators when designing system upgrades. Human factors design principles determine characteristics critical to maximizing task completion efficiency for human end users, but little work has directly addressed how such principles apply to working with robots. We lack the understanding of how designers create functional tool prototypes for non-human users, especially when they have limited knowledge about intended users.

Engineering and human factors design students designed a tool for a known (human), unknown (robot), or team operator (both). We performed content analysis on verbalizations accompanying early stages of tool design and final design presentations. Based on our results, we pose the need for *robot factors design* to determine what types and depth of information about robotic end users is necessary and sufficient during early stages of tool design.

This project was funded by the *Tufts Collaborates Seed Grant* program.

Table of Contents

List of Tables

List of Figures

Knowing End User Affects Design Process: Argument for Robot Factors Design

Imagine sorting large numbers of varied objects into predefined categories – such as cleaning up after you are done playing with a large box of Lego blocks. At first, you need time to attentively search for and categorize different pieces, and to memorize different types and shapes – therefore your sorting is slow. However with practice, you can sort almost effortlessly, having memorized the Lego types and their locations and you become more proficient in identifying them. Over time, you can reduce the time it takes you to clean up after you played. Human performance on tasks requiring specific skills improves with practice and repetition – the effect known as *the power law of practice* (Blackburn, 1936; Newell & Rosenbloom, 1981). We get better and faster at doing certain things as we practice, although the improvement rate diminishes over time as we become more proficient. But even with practice, cleaning up a Lego kit by hand is a time-consuming and frustrating activity – if there was a tool that could make sorting easier and faster, wouldn't you want to use it?

If you play only occasionally and with only one Lego kit at a time, having a tool to help you clean up would certainly be nice, but not critical to your enjoyment of Lego play time. But what if you need to sort not one large Lego box every now and then, but tens of such boxes on a daily basis? Schools and after school programs employing Lego building and design in classroom instruction face this problem on a large scale. A basic Lego NXT kit used in an educational setting contains approximately 16 types of Lego pieces varying in sizes and colors, totaling 430 pieces per kit. Even an experienced sorter needs a considerable amount of time to clean up a messy Lego NXT kit. In the best-case scenario, sorting multiple kits every day, regardless of sorter's proficiency, will lead to boredom and frustration. In worst case, the sorter will develop repetitive motion strain injuries. Manual sorting is therefore an extremely inefficient approach to maintaining Lego kits in working order. Prompted by such situations, designers develop tools or even automated systems to improve speed, easy, and task accuracy.

Human factors design practices emphasize the interaction between products (such as tools) or systems and their users. In essence, human factors professionals aim to design and optimize equipment to fit users' physical characteristics and cognitive abilities. Ideally, a tool developed to aid sorting a Lego kit could be adjusted to sorters' physical and mental capabilities – it should be easy to operate, fit the sorters' needs, and ultimately improve their satisfaction and efficiency. Understanding such functional relationship between humans and other system elements is supported by extensive research on principles, data, and methods optimizing both user's well-being and overall system performance (Wickens et al., 2004; Bevan & Macleod, 1994). Human factors design principles have been researched extensively (e.g., Maguire, 2001), mainly focusing on usability (although other factors, such as the aesthetic experience of interacting with tools and systems, have been considered; Proctor & Vu, 2016). Human factors design research focuses primarily on determining the end user characteristics and their relevance for system organization (Johnson, 2010). When designers conceptualize tools for other humans, they can approximate some end user characteristic by drawing on their own experiences. Such approximation is almost effortless and can improve speed and ease with which designers come up with new solutions. However, it can also fail to account for end user specific limitations that are not shared with the designer, resulting in tools or systems that do not pass usability tests. Relying on research-based approaches to end user capabilities is therefore a more cost-effective approach in the long run.

However, little work has directly addressed how end user characteristics – inferred or confirmed with data – are incorporated during the tool design process. This has led to a gap in our knowledge of human factors design: we seem to know what is best for our users, but we lack the understanding of how designers incorporate that knowledge in their work. The question of which cognitive mechanisms underlie the design process becomes particularly pertinent when we consider situations where the designer possesses little knowledge about the intended end user, yet is still charged with creating a functional prototype. How do designers, such as engineers and programmers, fill gaps in their knowledge about the end user during the design process?

The present study addressed this question by inviting engineering students to design a tool (a device) aiding mundane task completion for one of three end users: a typical person, a Baxter robot (Rethink Robotics, Inc.) with unknown technical specifications, and a team consisting of the two. We recorded our participants thinking out loud during a design session consisting of a 10-minute brainstorming portion followed by approximately 40 minutes of developing a final design and a short presentation explaining the tool idea to a hypothetical audience. We used these verbalizations to begin exploring approaches participants took to fill their knowledge gaps about the end user. Our analyses informed three questions:

- The process question: how do designers approach the design process?
- The assumption question: how do designers assume and explain information about their end users?
- The shared knowledge question: how do designers explain their designs to an uninformed audience?

Usability in human factors design

The term ‘user experience’ (UX) is associated with a wide variety of meanings (Forlizzi & Battarbee, 2004), ranging from how successful a particular tool is in fulfilling the users’ instrumental needs, all the way to subjective views on pleasantness and ease of use (Hassenzahl & Tractinsky, 2006). User-centered design focuses on collecting and analyzing what features and attributes a tool must have, or how it should preform, from the user's perspective (Courage & Baxter, 2005). Ignoring user requirements not only significantly decreases user satisfaction, but can result in financial losses due to redesign needs, increased maintenance costs, and possible workplace errors (Marcus, 2002).

A tool has increased usability to the extent that its features are logically organized, its use and components are easily learned, and it does not require prolonged attention for use. Usability is often affected by the tool's complexity, which in turn depends largely on the function it is meant to perform. Senders (2006) distinguishes between user, engineer, and technician complexity. *User complexity* implies the amount of effort required to operate a device; *engineer complexity* pertains to the level of the device's technological advancement; and *technician complexity* defines the scope of device's possible malfunction. For example, scissors – a handheld tool – are a device of low user complexity (the device is designed with the human hand's shape in mind and therefore a single demonstration is sufficient for learning how to operate it), low engineer complexity (two handles, two blades, and a screw), and low technician complexity (given very few possible malfunctions that can be alleviated with simple fixes, such as tightening the middle screw or sharpening the edges for more precise cutting). On the complete opposite end of the complexity spectrum are fixed automation solutions – systems where the order of operations is fixed by the equipment configuration – employed by large-scale factories, such as pulp and paper machinery that require extensive user training and highly specialized technical maintenance.

User experience is an umbrella term encompassing various measures that evaluate user complexity. For example, if you need to move large quantities of Lego blocks from one container to another, a small shovel-like tool is an excellent aid: it is easy and intuitive to operate even if one has never seen a shovel before, it improves task efficiency, and does not cause any discomfort with prolonged use. On the other hand, a pair of chopsticks is not a very user-friendly solution: it requires both instructions and practice to operate, it does not improve task efficiency, and prolonged use is guaranteed to result in both physical strain and frustration, lowering overall performance and satisfaction. User satisfaction is critical to operating tools, and therefore a good tool will account for intended user's limitations, skills, and preferences. Designers working with able-bodied individuals can refer to a preexisting mental model of what such individuals can and cannot do, but when the end user is less-than-typical, that mental model is no longer sufficient. This introduces a discrepancy between the assumed and the actual end user's capabilities through using an insufficient or faulty mental model. The resulting tool will therefore fail to fulfill the end user's instrumental needs, and will require expensive and time-consuming testing and redesign.

Our study investigated how designers approached user experience and usability when working with sufficient and insufficient mental models of their end users. When designing for a human confederate, participants could rely on their mental model of an able-bodied individual with no cognitive impairments. However, when designing for a robot, participants were given no information about their end user beyond its appearance, and as such needed to navigate the design process without sufficient background information. We restricted participants' access to end-user information in an attempt to explore the information designers consider critical to tool design – that is, what information the design process cannot proceed without. As such, our study is exploratory in nature and our findings should be generalized with caution.

Robotics development and conceptual challenges to UX

Fixed automation solutions are ideal for very large volume production, where highly specialized equipment proves to be the most cost effective approach. Production volume justifies high user, engineer, and technician complexity. However low and medium volume production does not warrant the expense of maintaining highly specialized, complex machinery, and call for more flexible, programmable solutions. Flexible and programmable automation is achieved by introducing adaptive robots, such as Baxter from Rethink Robotics, Inc. Baxter can be trained and programmed to perform a large variety of tasks.

The key advantage of solutions such as Baxter lie in its adaptability to the task at hand. The system is highly customizable and reprogrammable, and can be trained to complete tasks that require interaction with additional equipment for further flexibility. For example, a Baxter robot in a pharmaceutical company can complete a three-step process of opening a urine sample jar, pipetting an appropriate amount into a different jar, and placing the lid back on the sample jar. It is possible to equip one arm with a suction cup Baxter can use to hold onto the lid, and another arm with a gripper that serves a dual function of holding a jar in place and, later, operating a pipette. Automation of a repetitive task like this requires Baxter to interact with a tool – a pipette. A standard pipette will then need to be modified and its design optimized to allow for a successful interaction with a robotic end user whose capabilities differ from human motor skills. A pipette is likely a necessary tool for urine sampling whether the task is completed by Baxter or a human; at the same time, a human does not need the additional tool – a suction cup – that is crucial for Baxter's performance. The suction cup's usability is virtually nonexistent in human hand, and increases dramatically when the cup is utilized by a robot. Evaluating tool's usability is therefore heavily dependent on its potential users, emphasizing the need to consider user characteristics and preferences in tool design.

This poses a conceptual question in the context of user experience research: is Baxter a tool, or a tool user[1]?

Robots as tools and tool users

Whether robots such as Baxter are considered tools or tool users seems to be a simple conceptual question with little impact on real-life productivity. However how the robot is perceived by its operators or programmers has direct consequences for workload distribution, and therefore is directly linked to performance. Conceptualizing a robot as a complex tool requires its human operator to actively distribute workload (e.g., Durfee, Boerkoel, & Sleight, 2014), while a more versatile and autonomous robot that can be treated as a tool user allows workload to be allocated without explicit operator input. Paradoxically, having control authority over task allocation negatively impacts team fluency and significantly lowers user satisfaction and willingness to collaborate with a robot again (Gombolay, Gutierrez, Clarke, Sturla, & Shah, 2015). Given that increasing robot autonomy improves team performance and satisfaction, conceptualizing robots as tool users and not simply tools is the next logical step in improving productivity in contexts requiring collaboration between humans and robots.

How we perceive and stereotype robots has direct consequences for possible team dynamic. People depend on their preconceived notions about robots' reliability during interaction. Robinette and colleagues (2016) demonstrated that people tend to perceive robots as more knowledgeable and less prone to making mistakes, even after the robot has proven itself to be unreliable. Participants trusted a robot labeled “emergency guide robot” to lead them to safety in a mock fire scenario even after the robot failed to navigate successfully around the building during a non-emergency situation. This demonstration provides a stark contrast to a phenomenon called “algorithm aversion”: participants preferred to rely on forecasts of human forecasters, as opposed to a statistical algorithm, even if the algorithm outperformed a human forecaster (Dietvors, Simmons, & Massey, 2014). Coeckelbergh (2011) argues that we trust machines to which we delegate tasks, and a natural consequence of that attitude is trust in robots. The discrepancy between algorithm aversion and trust in robots can be driven by the degree of robots’ human-like appearance. Hinds, Roberts, and Jones (2004) argue that the experience of collaborating with or managing a robot is heavily influenced by whether the robot is machine-like or humanoid. People tend to retain more responsibility for successful task completion when working with machine-like robots as compared to humanoid robots, suggesting an inherently more collaborative approach to working with humanoid robots. As such, humanoid robots are generally seen more as tool users rather than tools.

How we conceptualize the robots’ role and their capability to perform a specific task influences how much we trust robots. Coeckelbergh points out that when a robot is considered an artifact, the trust is based on functional criteria – we trust our Roomba to vacuum the house, but not to transport items between rooms. However, if a robot is perceived to be an artificial agent, or more than a machine, the attitude towards it mimics our attitude towards animals or other people, therefore eliciting a certain amount of trust and expectations on its capabilities, such as the ability to make decisions or react to environmental cues. Capability judgments are heavily influenced by how human-like the robot appears: this effect is particularly pronounced when applied to androids, robots meant to look as lifelike as possible, but diminishes with interaction (Haring, Matsumoto, & Watanabe, 2013). We apply our perception of a human’s trustworthiness to robots when they are humanoid enough to warrant such transfer. Researchers argue that increasing human features in robots will make human-robot interaction easier by invoking rules for human social interaction and thus making robots more intuitive to operate (Brooks, 2002). How we perceive and stereotype robots depends on the machine type and interaction context, therefore the seemingly academic question of whether robots are tools or tool users has a very tangible impact on work dynamic in an environment involving robots.

In our study, designers working with human-robot team likely relied on their subjective perception of robots’ trustworthiness in accounting for human-robot work dynamic within their design.

Designing tools for robot users

Extensive human-robot interaction research contributed to our knowledge about how people treat and relate to robots in various contexts (Yanco & Drury, 2004). Based on such research, it is difficult to predict the breadth of stereotypes both designers and users rely on when filling in gaps in their knowledge on the robotic end user. Human-robot interactions are heavily influenced by what knowledge human operators may have about as specific robot, whether this knowledge is accurate or not (Hancock et al., 2011). Do such assumptions and stereotypes influence how designers approach working with robots as well? We approach this issue by introducing three different end users for whom our engineering students design tools. Data from participants designing for a human provided a basic overview of the design process practices necessary to solve this particular design problem. Designers working with human end users were able to rely on their preexisting knowledge about the end user’s capabilities. Such mental model was accessed easily and almost effortlessly, and therefore the main difficulty these participants faced was coming up with a tool solution. Data from participants designing for Baxter informed our understanding on how engineers fill in the knowledge gaps about the end user. In addition to coming up with a tool solution, these designers had to work with an insufficient mental model of end user capabilities. Lastly, we used data from participants designing for a team of a human and Baxter to infer how the stereotypes about a robot and a human end user change given how the designers conceptualize division of labor within the team. Human confederate’s presence in the team condition could lead to the robot being perceived more as a tool, given the proximity of a possible operator. On the other hand, the robot end user working on its own is lacking available operators, and as such could be conceptualized more as a tool user.

Design process as problem solving

User-centered design broadly relies on problem solving principles. Human problem solving in its most basic form requires generating possible solution paths to move between the initial problem state and the solution (Newell & Simon, 1972). The key aspect of reaching a successful solution is understanding the initial state, and clearly defining the desired outcome. Solvers can enhance their understanding of a complex problem by decomposing it into simpler sub-problems (Armstrong, Denniston, & Gordon, 1957). Breaking down the problem improves accuracy in coming up with an effective solution, and allows for a systematic evaluation of the user’s needs and capacities.

User-centered design identifies the need for a particular tool (initial state of the problem), analyzes user requirements (what features and attributes a final design should have, as well as how it should perform – the end state), and tests and readjusts possible solution paths through the process of iterative design (Courage & Baxter, 2005). The iterative design principle mandates that a product is designed, modified, and tested repeatedly until a desired outcome is reached. It consists of four phases: *concept*, *design*, *development*, and *release*. Concept and design phases are of particular importance to our work. They encompass how participants build the problem space specific to the tool design task, including evaluating and adjusting their preexisting mental models of end user capabilities. Development and release phases are not covered within our study design, as they focus on producing and testing prototypes, and therefore are mentioned only briefly. We elaborate on all four phases in the following section.

Concept phase

Courage and Baxter (2005) describe the concept phase of development as the "idea phase". The designer is conceptualizing usability and tool design goals and objectives, creating possible user profiles, and performing research on user requirements. Such research can involve fieldwork in form of interviews or field studies (Verhoef, 2007), or rely on designer's mental simulations. Applied Cognitive Task Analysis (ACTA; Militello & Hutton, 1998) is an approach in which user's cognitive processing requirements are mapped systematically to each step of a task. The designer begins with mentally simulating the task and interaction with the tool to delineate the basic steps. This initial design stage has the greatest relevance for human factors research, as it requires the designer to extract information about the cognitive demands and skills required for a task, and assess what kind of assistance the user will need to perform proficiently. This initial analysis and representation of the problem in turn influence the direction and range of proposed solutions, as well as the efficiency with which a solution is found (Reimann & Chi, 1989). We investigated the relationship between initial task analysis and proposed solutions by observing the designers' brainstorming (operationalized as the first 10 minutes of the design task) and comparing it to the presentation stage, where each participant explained their solution to a hypothetical audience. We investigated possible starting points, how they were moderated by a specific end user, and whether the end user affected problem analysis and interpretation.

Vessey (1991) points out the importance of what she calls 'cognitive fit': it is a cost-benefit characteristic suggesting that most efficient and effective problem solving occurs when the problem representation and any tools or aids available all support the strategies required to perform that task. Vessey and Galletta's (2001) work further discusses cognitive fit in the context of information visualization, where the data format plays an important role in aiding or disturbing inference making. It is nevertheless an important observation for tool design practices. If the information about end user's characteristics and preferences is not immediately available to the designer, or requires extensive additional processing to be converted into an easily accessible format, the resulting design will suffer. When designing a tool for human use, missing information can be temporarily supplemented by designer's personal experiences and inferences on average person's motor and cognitive skills. Such inferences are much harder to draw with respect to a robotic end user. Information about a robotic end user has to be drawn from another source that might not be immediately available – such as manuals, more experienced designers or robot operators, or simple interactions aimed at testing basic capabilities (Argote & Ingram, 2000). When such resources are unavailable, designers can create a knowledge base about the end user inferring capabilities from available evidence (for example, inferring video processing capabilities for robots with visible cameras), and supplement lacking information with a preexisting mental model of another end user.

To date, research has not explored how engineers conceptualize the needs and challenges in designing tools for robots. While the initial assessment of the problem and its mental representation are critical, engineers often cannot draw on their own experience when designing for robotic end users. We addressed this gap in literature by observing how the organization of the design process differs as a function of the end user (the process question). We posed a design problem to engineering students and stipulated a particular end user: a human, a robot, or a human-robot team. We provided little information on the end user's capabilities; the user was present during the design process and therefore certain capabilities could be inferred. In the human condition, a confederate sat across the table from the participant. In the robot condition, Baxter was positioned across the table from the participant, with one arm propped on the table, and the other arm hanging freely on Baxter's side. In the team condition, a confederate sat next to Baxter and both were positioned similarly to single human or single robot condition. Our participants could make certain inferences about the end users' capabilities via visual inspection – for example, they could see that the confederate was an average and seemingly able-bodied young female and could base their assumptions on their preexisting knowledge of an average person's capabilities. Similarly, they could see that the robot had a camera positioned next to its grippers, and could therefore assume certain capabilities based on video analysis algorithms. However, lack of direct interaction with the end user, paired with lack of additional information in form of a manual, forced designers to assume hypothetical capabilities.

We explored whether specific starting points to formalizing the problem space impact the overall productivity during the design session. In our study, participants spent approximately an hour coming up with potential tool solutions, and their work was not advanced to the prototyping or beta-testing stages. Overall design session productivity can be assessed with regard to the design outcome – such as the number of functional prototypes conceptualized. However, our analyses explored a limited subset of verbalizations (first ten minutes spent brainstorming possible solutions and a short presentation explaining the session's outcome). We defined *productivity* as the extent to which each participant focused on a possible design outcome (i.e., a tool). Looking at the starting point to designing a tool, did focusing on Lego kit's affordances as opposed to end user affordances result in a more productive design session? We also evaluated two possible influences the intended end user had on the design process: which starting points designers chose to begin with, and how designers interpreted task requirements.

Design phase

Information collected in the concept phase serves as a basis for iterative design creation. Potential tool users walk through a low-fidelity prototype (such as a pen and paper sketch) to ensure that all the user requirements are contained and possible to execute using a proposed solution; alternatively, designers mentally simulate how a tool user would interact with the prototype. Designers evaluate a number of heuristics they either employed during the concept phase or still need to employ at design phase. Our participants were not interacting with their end user (confederate or robot), therefore they were forced to mentally simulate users' interaction with the tool.

A simple starting point to filling knowledge gaps about user's capabilities is to use our own knowledge and experience (Nickerson, 2001). The projective way of knowing is an intuitively compelling approach, but can ultimately reinforce false assumptions and thus start the design process from an erroneous point. For example, experts have been shown to be particularly bad predictors of novice performance, even though presumably they have experiential knowledge of problems a novice will encounter (Hinds, 1999). This discrepancy can be attributed to a host of heuristics – cognitive “shortcuts” – that people employ in problem solving, and therefore commit logical errors in the design process. Designers may anchor their knowledge gaps on user’s capabilities with their own performance, and adjust as more information arises in the design and testing stages until a plausible estimate is reached – a phenomenon Epley and Gilovich (2006) termed *the anchoring heuristic.*

Anchoring heuristic and projective ways of knowing originate from a larger assertion that humans have mental models of their environment (Rouse & Morris, 1986). A mental model depicts only selected concepts that act as representations of real systems. Mental models are incomplete, as they do not represent the entire knowledge one possesses about a system, but rather a specialized subset of such knowledge (Forrester, 1971). Therefore, a user’s mental model reflects the understanding of a system, rather than the actual objective representation of said system (Carroll & Olson, 1987). Mental models inform mental simulations (Wilson, 2002) where the designer adapts the extant model to a new problem and design solution. Kankuzi and Sajaniemi (2016) point to inadequacy of mental models in usability-oriented design and advocate for data-driven approaches in design.

When a mental model proves to be insufficient, how do engineers infer information on the end user’s capabilities? We therefore posed the assumption question: how did designers infer and explain information about their end users? We compared how the end users’ capabilities were assessed and communicated during the design process depending on whether the participant was working with a single end user (human or robot) or a team. In our paradigm, engineers could make judgments on end user’s capabilities via visual inspection of the user’s appearance. We also allowed them to make assumptions on capabilities that they deemed necessary for design, but could not infer with certainty from visual inspection alone. We were interested in what kinds of assumptions engineers make when inferring end user's capabilities with regard to their mental model, and therefore what information designers consider critical to their work.

Advantages of using one's own mental models and cognitive heuristics in inferring end user abilities diminish when the designer is required to communicate with other designers. Successful communication requires designers to identify and appeal to *common ground* between a speaker and an addressee or audience – a set of shared knowledge, beliefs, and assumptions (Clark & Carlson, 1981; Clark & Marshall, 1981). Common ground is a mental model of shared knowledge, and it can encompass any range of knowledge, beliefs, assumptions, shared visual, auditory, or tactile perceptions, and even affective states and intuitions, as long as the referent can be readily and uniquely inferred from the current common ground (Clark & Bangerter, 2004). As such, common ground can be broadly defined as either *communal* (encompassing information common to members of specific communities, ranging from cultures to workplaces) or *personal* (encompassing information co-present to two people). When designers presented their tool designs to a hypothetical audience, they likely relied on communal common ground in understanding what the end user is or is not capable of doing.

Any knowledge transfer necessarily requires identification of common ground between speaker and recipient. Speakers differentiate types of information on the basis of what Halliday (1967) termed *information focus*. As such, speakers consider information either *given* (information speaker believes the listener already knows and considers true) or *new* (information speaker believes the listener does not yet know; Clark & Haviland, 1977). Following this distinction, designers presenting their tools elaborated or explained all information that they did not consider to be part of the common ground with the audience. Grice (1975, 1978) argued that such knowledge transfer will largely depend on the cooperative principle in language use, specifically - maxims of quantity and relevance. According to Grice, speakers are trying to cooperate with their addressees, and as such their presentations of end user capabilities should be relevant to the task and hand, and presented in an informative fashion but without overloading the addressee with unnecessary details.

When participants presented their tool designs and solutions to the sorting problem, they assumed an expert position vis a vis a novice audience. There is an inevitable difference in their knowledge about the tool or the problem solution itself. But how is this difference in knowledge manifested and addressed with regard to the intended end user?

Isaacs and Clark (1987) suggest that accommodating discrepancies in expertise can be divided into three processes: *assessing* expertise (experts directly confirming novices' expertise or novices demanding additional expert explanation), *supplying* expertise (experts supplementing complex claims with explanations without an explicit prompt), and *acquiring* expertise (novices making inferences from incidental information). Our participants presented their tools to a hypothetical audience, and assessing and acquiring expertise directly was not possible. Instead, participants had to rely on their subjective assessment of what a hypothetical, educated and uninformed audience member knew about the intended end user, and supply relevant expertise accordingly. Presenting to a hypothetical audience may therefore differ from presenting to an actual audience. Schober (1993) investigated how perspective choices in giving directions change when participants convey spatial directions to a hypothetical audience and a speaking partner. People conveying spatial descriptions rely either on their egocentric position (deitic use) or canonical object orientation (intrinsic use). Schober observed that solo-speakers were almost uniform in being non-egocentric, while speakers with partners varied in what positions they assumed. When unable to assess and acquire expertise directly from a conversational partner, speakers adjusted their explanations accordingly to ensure that the supplied expertise was accessible to any possible listener. As such, Schober’s participants knew that their hypothetical audience would not be able to see what objects the participants were pointing to, and therefore they had no common ground justifying the use of egocentric perspective directions.

Our study’s setting allowed us to pose the shared knowledge question: how do designers explain their designs? Our participants needed to first assess the expertise of a hypothetical audience and supply new information accordingly, therefore information that was crucial to their presentation but not verbalized belonged to the assumed common ground. We examined whether participants explicitly discussed end user capabilities and directly assigned certain tasks and responsibilities to a specific end user. We theorized that explicit statements address gaps in common ground, while omitting certain information about the end user – that is simply not verbalizing it – implied that such information was within common ground. We also explored whether the intended end user had an impact on how much time participants needed to fully explain their solutions, which reflected the extent to which information pertinent to the tool solution was not encompassed by the common ground and needed to be reiterated.

Development and release phases

Once a satisfactory model is designed during the design phase, designers proceed to develop pre-product prototypes. Prototypes lend themselves to additional testing, where designers simulate real-world conditions under which the actual product will be implemented. Additional beta testing can be performed, when the product is released to a limited audience for further feedback. Pending any adjustments warranted by the testing results, the product is released and accompanied by follow up studies (such as surveys or interviews) with real users, and often site visits to evaluate how the product fulfills its purpose once it is being used in its environment. In this study, we asked participants to conceptualize a possible tool solution given a specific end user, however the task ends with participants explaining their ideas only after a relatively short design session. We were interested in how participants defined their initial problem space and assessed end user capabilities, therefore advancing the tools designed during the study to the development and release phases was beyond the scope of our research questions.

Design problem/Current study

In the current study, we were interested in observing the concept and design phases of product design. We posed the following design problem to a group of engineering students: Lego Mindstorms NXT series is a kit containing different types of Lego pieces, totaling over 430 pieces distributed across two trays with 4 and 13 compartments each. Sorting a messy NXT kit is a mundane, time-consuming task. The goal of the design problem was to optimize the sorting process by conceptualizing a tool (a physical piece of equipment) to be used by a particular end user in sorting a single NXT kit according to instructions.

Each participant designed a tool for one of three possible end users:

- A robot
- A human
- A robot-human team

First, participants brainstormed possible solutions for about 10 minutes while thinking aloud. Then they were given approximately 40 minutes to sketch out their idea. Lastly, participants presented and explained their tool concept to a hypothetical audience.

The current work’s goal was to observe whether – and if yes, how – the end user of the potential tool impacted the design process practices employed by engineers. We focused on the initial 10-minute brainstorm, which encompassed participants setting up the problem space for their design session, as well as identifying and addressing possible gaps in their end user mental models. We then explored how participants presented their designs to a hypothetical audience, which indicated what kind of information participants considered to be common ground. We explored three aspects of the design process:

- Process question: how do designers approach the design process?
- Assumption question: how do designers assume and explain information about their end users?
- Shared knowledge question: how do designers explain their designs?

It is important to note the exploratory nature of this work. Results presented in this paper reflect initial analyses conducted on a limited subset of data (think aloud recordings from beginning and end of each session).

Method

Participants

Fifty (49 native speakers of English, 27 female) undergraduate and pre-Master’s students (age M=19.7, SD=2.5) from the Tufts University School of Engineering completed the study for monetary compensation. Five participants were excluded from the final analysis due to difficulties comprehending the task (3 participants due to reluctance to think aloud, 1 due to lack of English proficiency) and recording equipment malfunction (1 participant). Participants came from engineering (mechanical, civil, biomedical, environmental, electrical, and general), human factors, and computer science departments. Recruitment aimed to find inexperienced, yet informed and motivated designers. Participants were randomly assigned to one of three end user conditions. 15 participants per condition (24 female) are included in the final analysis, with one participant from Team scenario missing their presentation phase due to equipment malfunction.

Procedure

Verbal protocol. Think aloud training used a procedure previously used by Tenbrink and Taylor (2015):

> I will, in a minute, give you a task to perform. While you do that, I will ask you to THINK ALOUD during the whole procedure of the task. We are interested in what you think about as you perform the task. Therefore I want you to say EVERYTHING you are thinking from start to finish of the task. Don't try to plan out what you say and don't talk to ME. Just act as if you were speaking to yourself. It is most important that you keep talking, even though you won't get any response or feedback. Do you understand what I want you to do? If I do not hear you talking for a bit, I will remind you that you are to say aloud what you are thinking. Good, now we will begin with some practice problems. First, I want you to multiply two numbers in your head and speak out loud what you are thinking as you get an answer. What is the result of multiplying 24 x 36? Good. Any questions? Here's your next practice problem: How many windows are there in a house you used to live in—for example your parents' house?

End user. Participants designed a tool for one of the three end users:

1) Single robot –Baxter (Rethink Robotics, Inc.) served as the robot end-user. The robot was positioned across the table from the participant, and was switched off. One arm was propped on the table to allow participants a detailed view, while the other arm hung in the off position behind the table. No additional information about the robot's capabilities or technical specifications was provided, and participants did not have any experience working with Baxter prior to participation. Participants did not interact with the robot, and were discouraged from touching it. Throughout the study, the robot was referred to as either "Baxter" or "the robot."

2) Single human – a female confederate (one of three researchers) sat across the table from the participant. Participants had no personal connection to the confederate. The confederate did not interact with participants in any way, and avoided looking directly at their work. Throughout the study, the confederate was referred to by her first name.

3) A robot-human team – a female confederate was seated alongside Baxter. Robot's setup and confederate's instructions matched the previous two conditions.

Design problem. Participants received a mixed up Lego^{TM} NXT kit with a printed picture of what an ordered kit looks like. They received the following instructions:

> Are you familiar with these? This is a Lego NXT kit. It is used widely in engineering education. We bring students into a classroom where they use an NXT kit to complete a certain task. After they are done with their assignment, the kit needs to be sorted before it can be given to another student. The kit in front of you is a sample mixed up kit, and the picture shows what it looks like when it is sorted. As you can tell, sorting a mixed up kit to resemble the one in the picture is a mundane, time consuming task. In this study, I ask you to come up with a tool that will help the agent in front of you: (Baxter, person, team consisting of Baxter and person) sort NXT kits more efficiently. Do you have any questions?
>
> I understand you may not be familiar with the full extent of (Baxter's/person's) capabilities. If a particular capability or feature is crucial for your design, such as for example whether (Baxter, person) is 7 feet tall, you can make that assumption. Just make sure to state out loud what you are assuming about (Baxter's/ persons') capabilities.

Task materials and recording. Participants initially received a sample mixed up Lego^{TM} NXT kit and a reference picture of what a sorted kit looks like. Participants received plain paper, scissors, pens/pencils, and tape they could use to develop a model of their tool idea. The verbal protocol was recorded using the microphone built into SMI Mobile Eye Tracking Glasses (SensoMotoric Instruments, Inc., Germany). The eye movement analysis is not included in this paper, therefore all information regarding the eye tracking recordings is omitted.

Procedure. Participants were fully informed about the task procedure and trained to think aloud. They were seated comfortably at a table and asked to remain seated during the study. The intended end user of the tool was positioned on the other side of the table, and participants could not touch or otherwise interact with the end user. With the robotic user, participants had little prior knowledge of technical specifications or experience working with the robot.

Participants received a messy Lego™ NXT kit that they could freely manipulate. The design task consisted of three stages: first, participants brainstormed possible solutions for 10 minutes while thinking aloud. Then they could use paper, pencils, scissors, and tape, to record their design idea. After approximately 40 minutes, participants gave a short presentation of how their tool can be used to alleviate some efficiency issues inherent to the sorting process. Participants received the following instructions:

> First, you will be given a 10-minute brainstorming session. During this session, I ask you to think aloud as you are doing precisely that: brainstorm. You may manipulate the Lego pieces as you wish. Please remain seated throughout the task. Lastly, your intended agent – (Baxter, person, team consisting of Baxter and a person) is in front of you. (Please refrain from touching or moving Baxter in any way. I understand that the ability to touch the robot you are designing a tool for is important for this task, but one of the experimental limitations is that you refrain from doing that/The person will not respond to you or interact with you in any way).
>
> After 10 minutes, I will give you scissors, paper, pencils, and tape, and I will ask you to record your design idea in whatever way you want using the materials in front of you. I would like you to remember to think aloud as you do this. You may continue brainstorming as you record your idea. Lastly, when you're done recording your idea, I would like you to explain it as if you were presenting to an educated audience that is unfamiliar with (Baxter's/ person's skills and) the task of sorting NXT kits. I will remind you of this again when your brainstorming and design time is over. You will have to explain the problem you are solving, and your solution.
>
> I will be sitting behind you. You have 10 minutes to brainstorm. Please remember to think out loud as you brainstorm. You will not hear any

> feedback from me. I will prompt you to keep on talking if you remain silent for some time. Also if you have any questions or doubts, you can ask me and I will answer. You can start whenever you are ready.

Lastly, participants completed a short demographic questionnaire recording their age, gender, English fluency, major, and prior experience relevant to the tool design task.

Analysis

Research questions

We explored the initial 10 minutes brainstorming session, and the presentation following the design session. Our designers set up the problem space during the brainstorming session, which allowed us to explore their preexisting knowledge of end user, as well as observe preferred problem solving strategies. The presentation session was geared towards summarizing and communicating results to a hypothetical, educated but uninformed educated audience, and as such reflected how designers conceptualized the common ground with their audience.

We explored three questions with associated sub-questions:

- **Process question**: how did designers approach the design process?
    - What was the most productive starting point to designing a tool? We investigated whether exploring Lego kit affordances or end user affordances corresponded to how much participants focused on their design ideas, as opposed to problematic issues inherent to the sorting task.
    - Was the starting point moderated by the intended end user? We compared whether participants working with different end users favored exploring Lego kit affordances over end user affordances during the brainstorming session.

  - Did participants think differently about the task requirements depending on their end user? We explored whether participants interpreted the task problem space differently depending on their end user during the brainstorming session.
- **Assumption question**: how did designers assume and explain information about their end users?
  - Was there a difference in how designers communicate end user requirements? We investigated whether participants addressed the end user's role in the sorting task explicitly or not both during the brainstorming and presentation sessions.
  - How did designers think about their end user's requirements? We investigated whether end users' capabilities and limitations were directly discussed and possibly explained during the brainstorming session.
  - Were designers critical of their end users? We compared how frequently participants focused on the end user's limitations and their influence on the ability to complete the sorting process, as opposed to evaluating whether a particular tool or sorting strategy was feasible given end user's strengths.
- **Shared knowledge question**: how did designers explain their designs?
  - Did designers present their solutions differently depending on end user? We compared presentation length and focus on end user involvement in the sorting task and tool operation during the presentation session to assess

what information the designers considered to be common knowledge (shared knowledge with a hypothetical audience).

Cognitive Discourse Analysis (CODA)

Our analysis employed Cognitive Discourse Analysis (Tenbrink, 2014). CODA can be used to investigate verbalizations of cognitive processing, including processes such as problem solving and cognitive strategies and heuristics, making it a fitting approach both to the form of data and research questions of interest to this project. The central idea of CODA is that unconstrained natural language elicited in purposefully controlled situations is a rich data source ideally to be combined with other modalities or representations of cognitive processes. This approach can be combined with other data modalities, such as eye tracking, or serve as a standalone data source.

The analytic approached outlined by Tenbrink (2014) emphasizes two approaches to verbal data: content-based inspection (particularly when carried out by experts in the problem domain), and analysis of linguistic features (reflecting conceptual phenomena of interest). Our analysis focused heavily on content-based inspection.

Data preparation

Transcription and segmentation. Recorded think aloud data was transcribed verbatim using Infinity USB digital foot controller and Express Scribe transcription software (NCH Software, Inc.). A different researcher proofread each transcript to ensure accuracy. Seven coders participated in transcription and proofreading.

Transcripts were segmented into smaller units serving as the basis of analysis. The unitization was executed based on “content units.” Given that the participants thought out loud during brainstorming, it was difficult to parse the data based on grammatical structures such as sentences, as many sentences produced by our participants were complex and often did not have a defined end.

Therefore, we defined a segment or a unit as the shortest possible logical sentence-like utterance containing a single claim about the end user, the design, or possible references playing into it. For example, the following fragment was separated into five segments (denoted with “/”): *in my experience the most annoying thing about legos, something like that, sorting through them is just picking up the smallest pieces, / so if you can just pick up everything as a whole, / and then from there separate out the smaller pieces, / that makes it easier, / because then you’re just picking smaller pieces from smaller pieces and not from other bigger pieces*.

This approach to segmentation was tailored towards content analysis of the think aloud protocols, yet still allowed for extrapolation of linguistic features and patterns associated with particular concepts emergent in the protocol.

Similar to the process of transcribing, the transcript segmentation was proofread by another researcher. Four researchers were involved in segmenting; discrepancies between initial segmentation and proofread segmentation were resolved by a third researcher or discussed among the two and resolved mutually.

Content analysis. We categorized each utterance as belonging to one of five categories, with partial categories when appropriate:

- *End user*: exploring preexisting or standalone (i.e., not required by the specific context or design issue) capabilities. This category indicated participants building a knowledge base critical for further tool development.
- *Lego kit:* assessing content, composition, organization, or any other standalone characteristic of the Lego kit. This category allowed participants to explore the affordances of the sorting task at hand.
- *Design:* any utterances pertaining to what the final design could, should, or will look like, including all aspects of the tool and operation, as well as tool-independent sorting strategies. This category encompassed possible tool solutions, as well as how participants conceptualized the intended end user's role in the sorting process.
- *Process/Evaluation/Problem:* evaluating one's thinking, actions, and possible difficulties or limitations of a specific sorting solution. This category demonstrated thinking about and questioning design ideas as well as doubting or reflecting upon specific end user capabilities that impact the plausibility of the designed tool.
- *Task clarification and reiteration*: any utterance clarifying the task, including dialogue with experimenter. This category reflected participants' understanding of the experimental task.

The analyses did not include utterances from the task clarification and reiteration category, as these utterances contributed to understanding of the experimental task, and not solving the tool design problem. Partial category details are illustrated in detail in Table 1.

Several partial category items were particularly pertinent to our analysis:

- End user:
    - *Assumption* – explicitly verbalized possible capacities of the end user, not directly required by the tool idea. This sub-category reflected participants spontaneously inferring end user capabilities.
    - *Limitation* – explicitly verbalized lack of possible capacities of the end user, not directly required by the tool idea. This sub-category reflected participants spontaneously inferring end user limitations.
    - *Explanation* – explanation for why a certain assumption/limitation was inferred.
- Design:
    - *Action* - explicitly verbalizing an action that a specific end user needed to perform in order to operate the tool and/or contribute to the sorting process. This sub-category reflected participants designating end user tasks specifically in the context of tool operation.
    - *Strategy* - verbalizing an action needed to aid the sorting process without specifically addressing the end user. This sub-category reflected participants discussing a sorting strategy or actions needed to be taken without naming specific end user's involvement.
    - *Capability* - verbalizing what kind of capability was crucial for successful tool usage or participation in sorting. This sub-category reflected participants inferring or stating end user capabilities necessary specifically for operating a tool.

    - *Feature* – a characteristic of Lego kit, hypothesized or observed, that was critical to tool's operation or functionality. This sub-category reflected participants engaging specific task affordances in their tool design.
- Process/Evaluation/Problem:
    - *Question idea* – deliberating whether an idea or a tool was plausible and could function despite certain limitations.
    - *Question user* – deliberating whether the indented end user was capable of operating the tool despite certain limitations.

All utterances were categorized, and all categories and partial categories were mutually exclusive. Five researchers were involved in coding. Discrepancies in initial coding and proofread coding were resolved either by mutual discussion or by a third researcher.

We tested a random subset of analyses and achieved inter-coder reliability of .88. Following Krippendorff's (2004) assessment, obtaining reliability of $\alpha \geq .800$ is considered a reliable analysis.

Results

Variable types

We extracted three variable types from coded verbal protocol transcripts:

- Frequency: ratio of utterances categorized as a specific category to all utterances spoken by the participant within brainstorming or presentation. This variable assessed the extent of participant's focus on a specific category relative to the rest of the task.

- Binary: whether a category was utilized by the participant or not during brainstorming or presentation.
- Time to completion: how much time each participant spent focusing on a specific aspect of the task. Specifically, we looked at how long each participant spent explaining their final tool design.

Process questions

What is the most productive starting point to designing a tool? For this analysis, we defined *productivity* as extent to which participants focused on final tool ideas during the task. First, we explored the tradeoffs and benefits of choosing a specific starting point. Designers could begin either by examining the affordances of the sorting task (here approximated by exploring the affordances of the Lego kit), or the affordances of their specific end user. As their end goal was to come up with an idea for a sorting tool, we pose that higher frequency of utterances categorized as *design* indicates higher productivity during brainstorming. Similarly, higher frequency of utterances categorized as *product/evaluation/problem* indicates elaboration and, during brainstorming, hesitation to accept an idea. Whereas thinking about limitations contributes to the final tool design and as such contributes to designer’s overall productivity, this distinction allowed us to compare two separate aspects of tool design: focusing on the task’s end goal, and the underlying cognitive processes by which the end goal was conceptualized.

We demonstrated that during brainstorming end user utterance frequency correlated with design utterance frequency ($r = .301$, $p = .045$, Fig. 1), however end user utterance frequency did not significantly correlate with product/evaluation/problem utterance frequency. On the other hand, Lego kit utterance frequency corresponded negatively both with design frequency (r: = -.513, p < .001, Fig. 2) and product/evaluation/problem frequency (r = -.301, p = .017, Fig. 3).

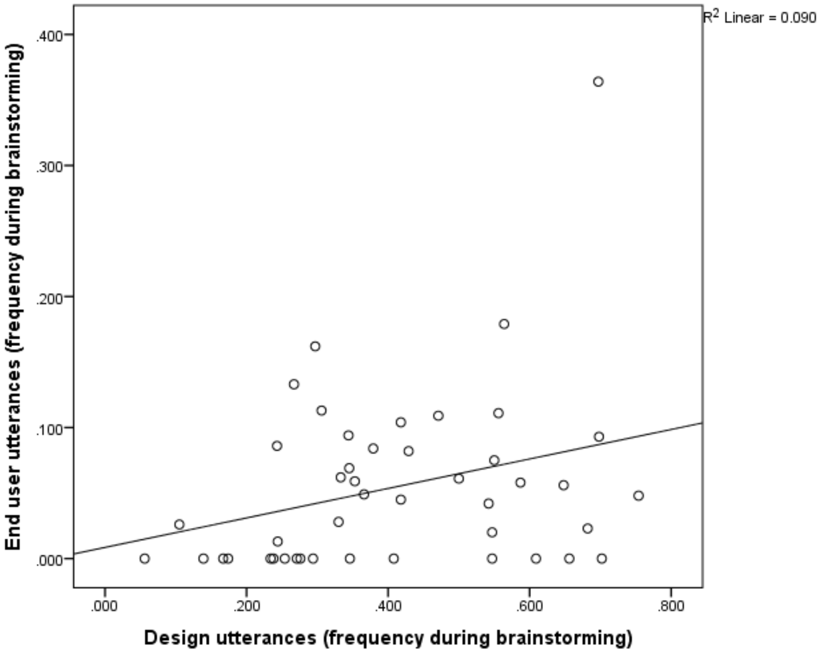

Figure 1. Correlation between end user frequency and design frequency during brainstorming.

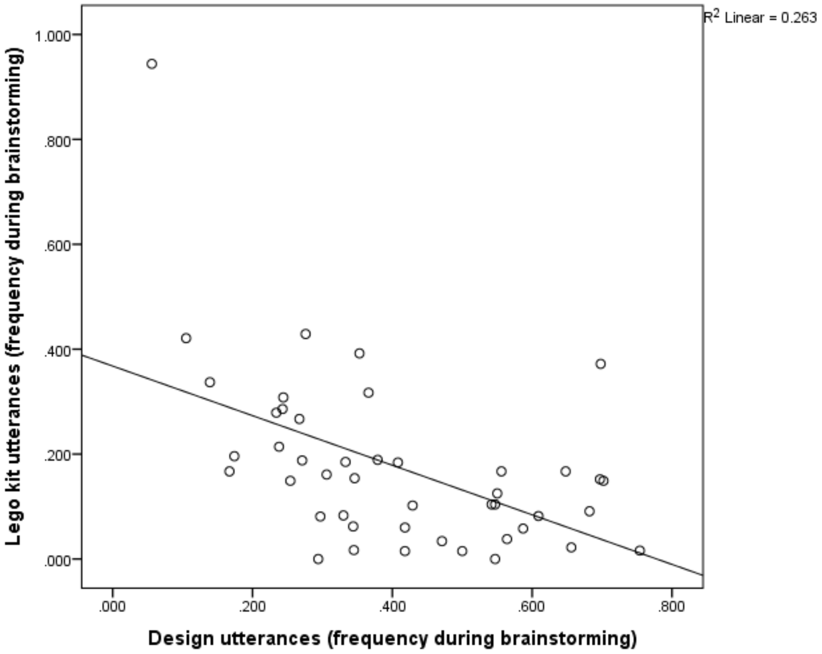

Figure 2. Correlation between Lego kit frequency and design frequency during brainstorming.

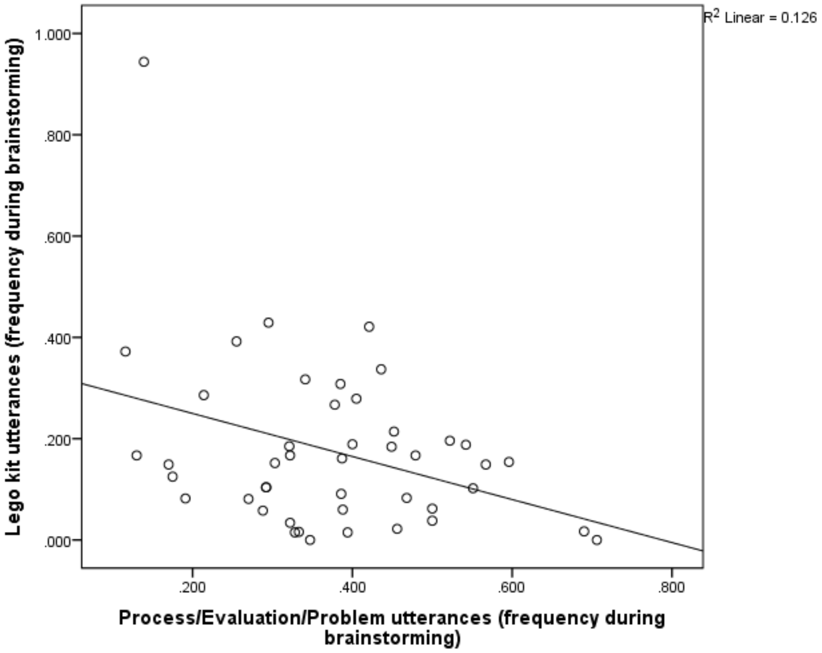

Figure 3. Correlation between Lego kit frequency and process/evaluation/problem frequency during brainstorming.

However these correlations did not carry over to the presentation phase. None of the relationships between the starting point (Lego kit or end user) and session productivity (focus on design or product/evaluation/problem) emerged for the presentation stage. It appears that the initial disadvantage of starting with Lego kit exploration did not affect the final product. This suggests that exploring Lego kit's affordances could be a more difficult starting point, but the advantages of first focusing primarily on the end user are short lived.

Is the starting point moderated by the intended end user? When comparing intended end user groups, it appears that the tendency to focus on Lego kit affordances was specific to designers working with a human end user. A repeated measures ANOVA comparing end user utterance frequency with Lego kit utterance frequency during the brainstorming session (F(2,42) = 4.894, p = .012) demonstrated that designers working with the robot and the team tended to split their attention between Lego affordances and end user affordances, while designers working with the human end user almost exclusively focused on the Lego kit affordances, failing to specifically address the end user (Fig. 4).

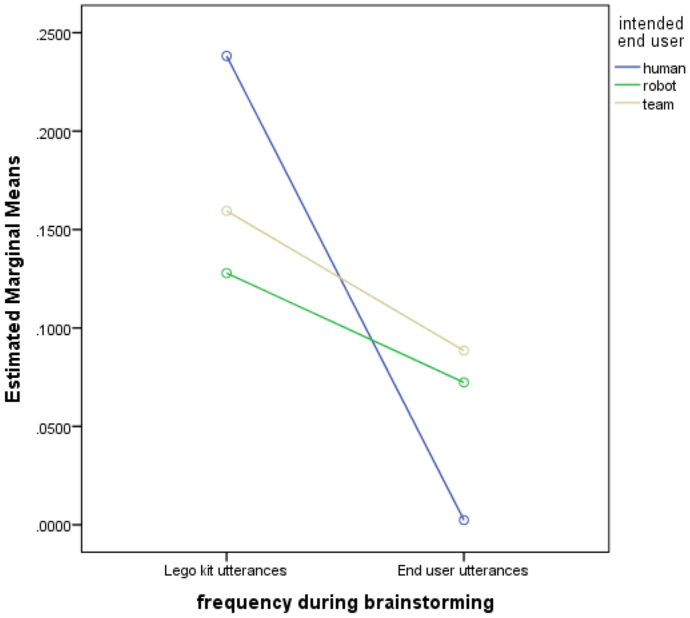

Figure 4. Repeated measures ANOVA comparing Lego frequency and end user frequency across different end users during the brainstorm.

Do participants think differently about the task requirements depending on their end user? We conducted a Chi-square test to evaluate whether incorporating Lego kit affordances in the design process was moderated by the intended end user. We demonstrated that participants did not significantly differ in referring to Lego kit "features" (a specific Lego affordance crucial to successful sorting, Table 1) during the brainstorm, (χ^2(2) = .297, p = .862). Lego kit played an important role in developing tools for sorting regardless of the intended end user.

We conducted a Chi-square test comparing whether designers conceptualized new ideas, or referred to existing tools in conceptualizing their solution (such as coin sorters or lettuce spinners). There was no difference in deployment of new versus inspired ideas across intended end users (χ^2 (2) = 1.996, p = .369). These results suggest that between-group differences in how designers approach their end users are not driven by individual designer preferences or specific design task characteristics.

Assumption questions

Is there a difference in how designers communicate end user requirements? We classified explicit statements (verbalizations) of end user involvement in the sorting process under *action* category, while *strategy* pertained to any utterance where a designer would indicate that something must be done in order to sort the kit – but would not directly address or implicate the end user. For example, an utterance was categorized as *action* when the end user was directly implicated in the task: “Baxter will put this piece here”. On the other hand, an utterance was categorized as *strategy* when a task needed to be completed but the participant did not specify by whom, such as: “this piece needs to be put here”. Designers used these simple differences in phrasing depending on having a robot or human end user. The action category was used significantly more with regard to the robot as compared to human (t(28) = -2.998, p = .006), while the strategy category was used significantly more with regard to the human as compared to robot (t(28) = 2.158, p = .040).

If we assumed that *strategy* utterances were meant to pertain directly to the human end user due to the ambiguous use of language and merge *strategy* and *action* sub-categories, an independent samples t-test demonstrated no difference between how frequently designers incorporated their end user in their design, both during brainstorming (t(28) = -1.404, p = .171) and the final presentation ((t(28) = -1.613, p = .118). This finding suggests that both end users played a crucial role in the design process, but the knowledge about the robotic end user was less established and therefore was verbalized – as opposed to the assumed knowledge of human end user capacities. This tendency persisted when we compared how the robot and human are addressed individually (and not as a team) in the team condition: a paired samples t-test demonstrated no difference in frequency of addressing either team member during brainstorm (t(14) = .077, p = .940) or presentation (t(14) = .422, p = .68).

When we looked at the frequency of direct end user affordance assessments (that is, all utterances falling under the *end user* category: referring to standalone end user capabilities), during the brainstorm almost no attention was paid to the human end user as compared to robot or team (one way ANOVA: F(2,42) = 9.560, p < .001, Fig. 5).

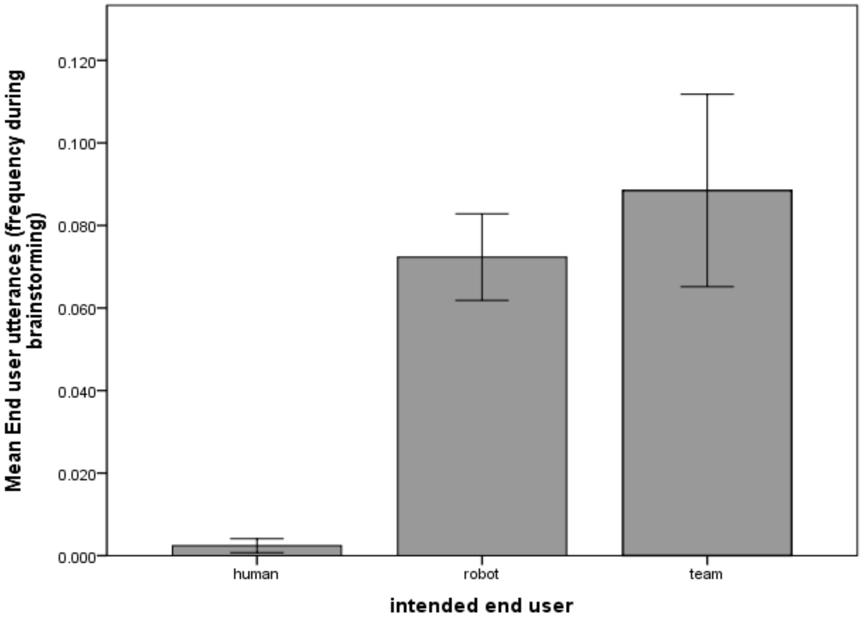

Figure 5. One way ANOVA comparing frequency of end user-related utterances across conditions. Error bars represent +/- 1 standard error.

How do designers think about their end users' requirements? We explored what kinds of approaches our designers employed in thinking about their end users. We conducted a series of Chi square tests assessing whether certain end users prompted designers to employ specific cognitive schemata in thinking about solutions.

Designers working with robots and teams relied heavily on verbalizing specific assumptions of end user capabilities, as opposed to designers working with humans who rarely make explicit assumptions: χ^2 (2) = 25.797, p < .001 (Fig. 6). Furthermore, designers working with robot and team end users addressed end user limitations significantly more as compared to designers working with human end users: χ^2 (2) = 9.514, p = .009 (Fig. 7). Lastly, designers working with teams/robots provided more explanations about their end users' capabilities and limitations, as opposed to designers working with humans: χ^2 (2) = 6.429, p = .040 (Fig. 8).

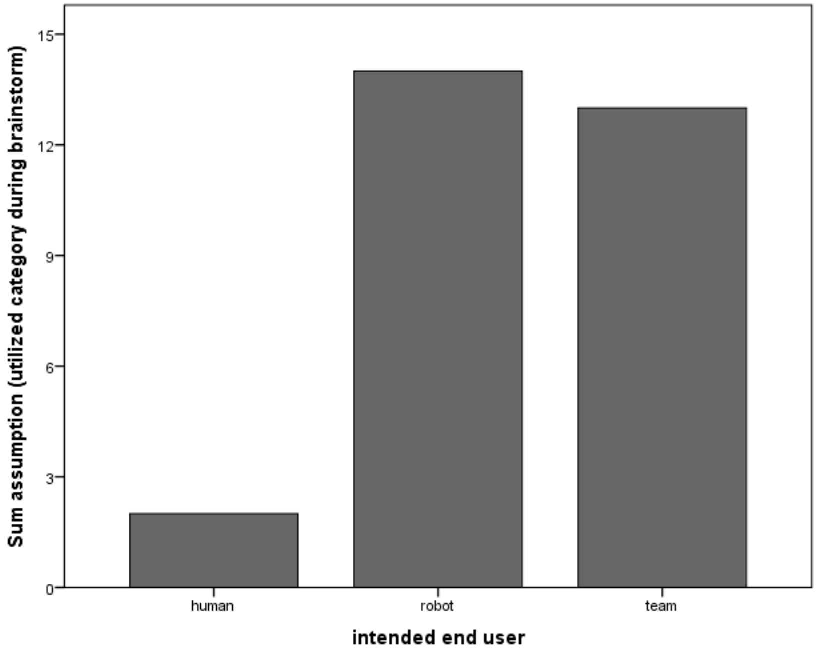

Figure 6. Total number of participants making end user assumptions across groups during brainstorm (*assumption* code).

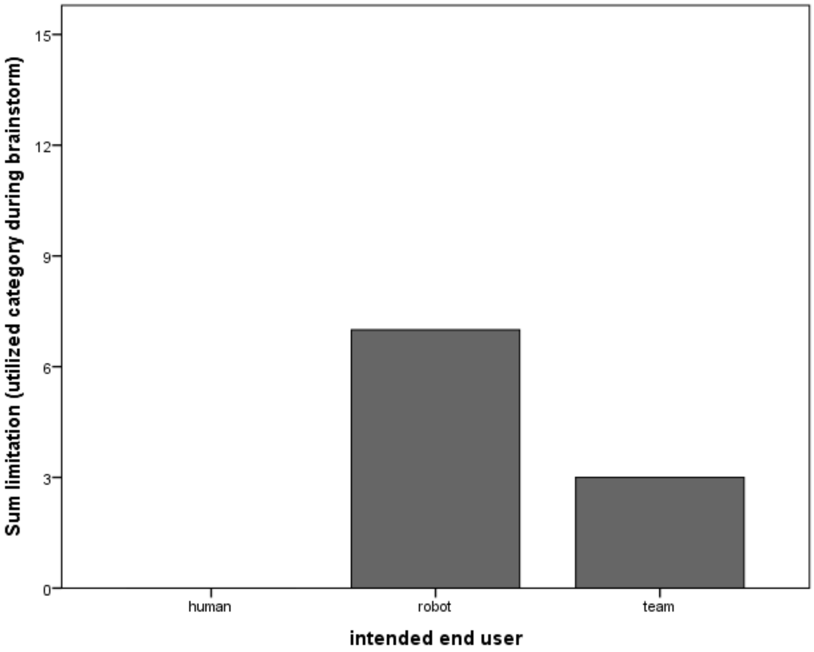

Figure 7. Total number of participants commenting on end user limitations across groups during brainstorm (*limitation* code).

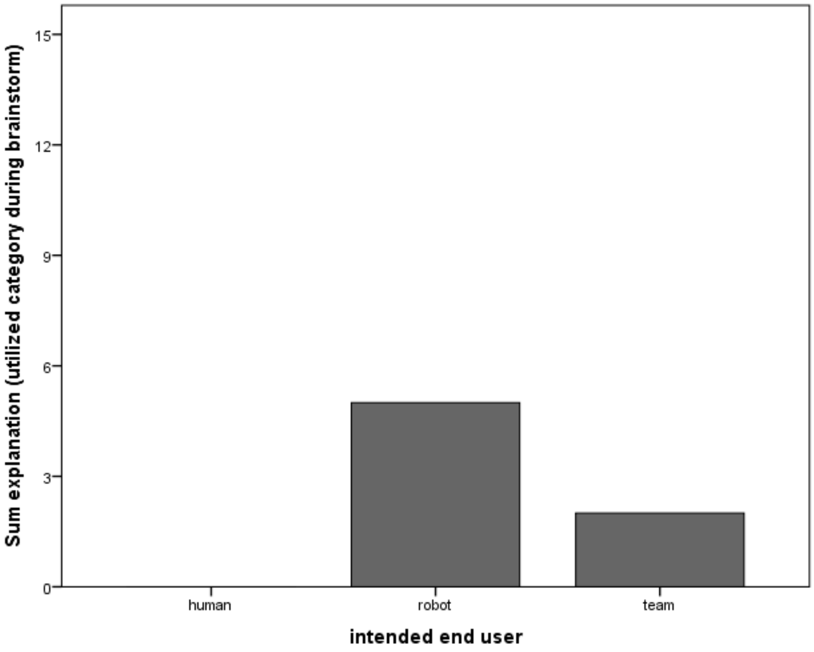

Figure 8. Total number of participants explaining end user comments across groups during brainstorm (*explanation* code).

Are designers critical of their end users? Designers working with humans never questioned whether their end users could operate the newly conceptualized tool, as opposed to designers working with robots/teams (χ^2 (2) = 9.303, p = .010, Fig. 10).

Designers working with human end users questioned whether their design ideas were possible during the brainstorming session more explicitly than designers working with robots (χ^2 (2) = 8.927, p = .012, Fig. 9). This pattern indicates that designers working with humans were not simply less critical of their work, but perhaps did not need to question their intended end user's capabilities.

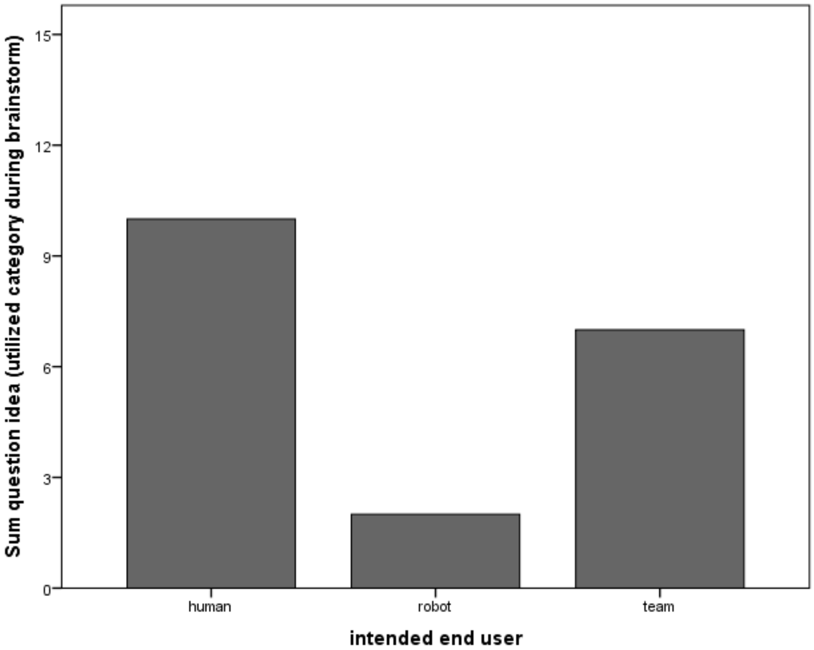

Figure 9. Total number of participants questioning their ideas during brainstorming session, across groups (*question idea* code).

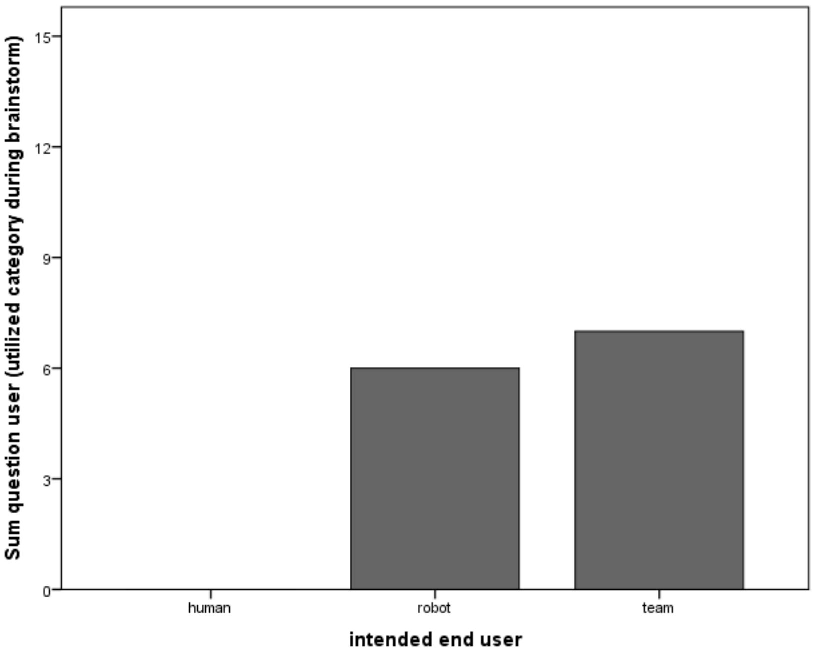

Figure 10. Total number of participants questioning their users during brainstorming session, across groups (*question user* code).

Shared knowledge question

Do designers present their solutions differently based on their end users?

Participants working with robots and teams required significantly more time to explain their ideas, as compared to designers working with humans (F(2,41)=5.384, p=.008, Fig. 11). To rule out individual differences in speaking speed, we compared participants' words per minute rates during their tool presentations, and found no significant differences in words spoken per minute across end users. This confirms that the difference in presentation duration was not related to slower speech or more frequent long pauses in speech.

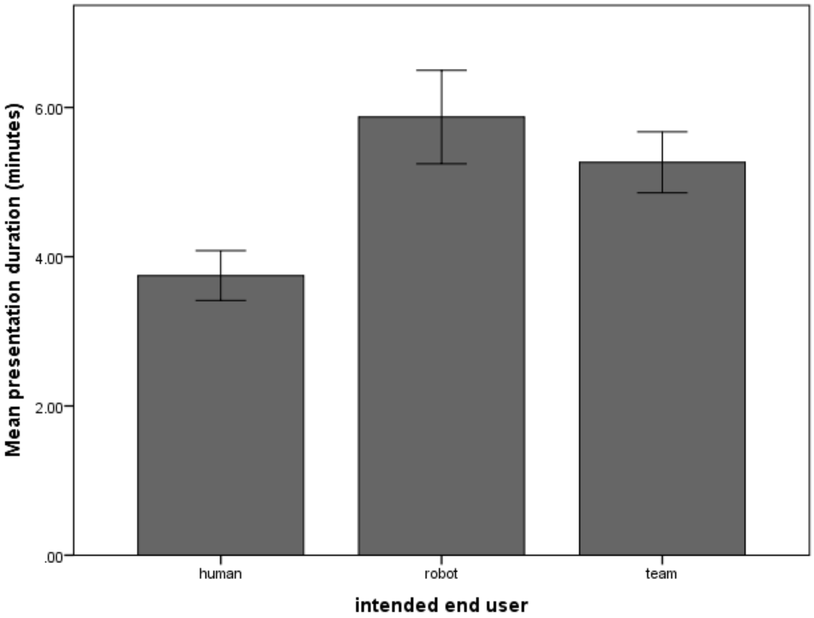

Figure 11. Mean presentation duration (in minutes). Error bars represent +/- 1 standard error.

Designers differed in their approaches to explaining the end user's involvement in the sorting task. Robotic and team users were directly implicated in the sorting task (*action* code) as opposed to human end users (χ^2 (2) = 24.210, p < .001, Fig. 12), while overall sorting strategies without direct end user involvement (*strategy* code) were mostly discussed in presentations for human and team end users (χ^2 (2) = 13.821, p = .001, Fig. 13). Human capabilities were also questioned less explicitly than those of a robot or team (χ^2 (2) = 13.403, p = .001, Fig. 14).

However we found no significant differences in how designers discussed end user limitations (*limitation* code) across conditions (χ^2 (2) = 1.224, p = .542).

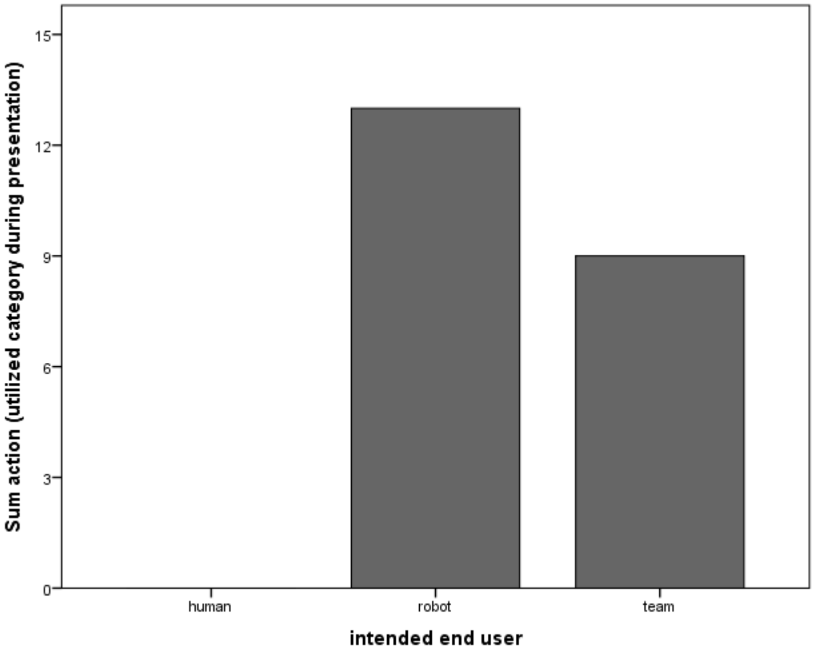

Figure 12. Total number of participants emphasizing direct end user involvement in the final tool presentation (*action* code).

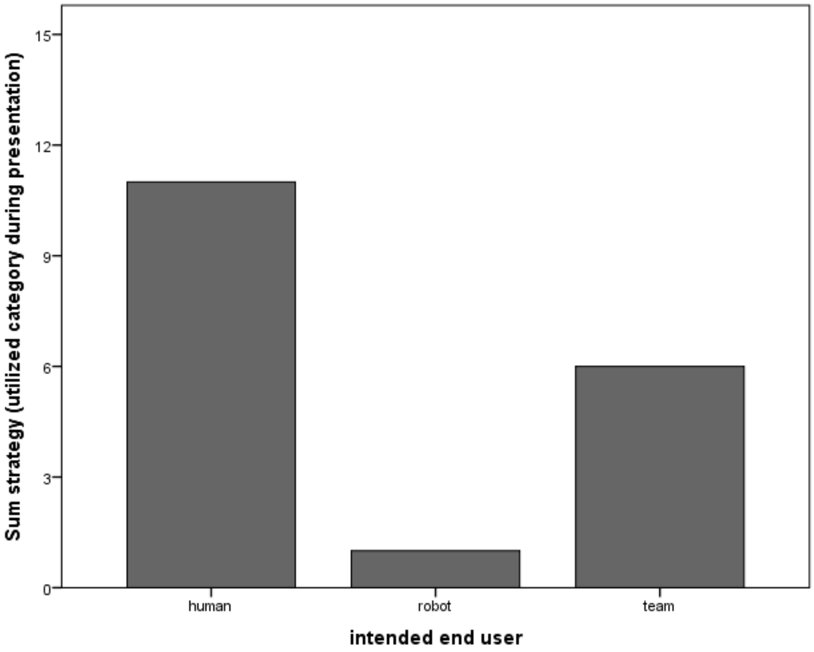

Figure 13. Total number of participants emphasizing indirect end user involvement in the final tool presentation (*strategy* code).

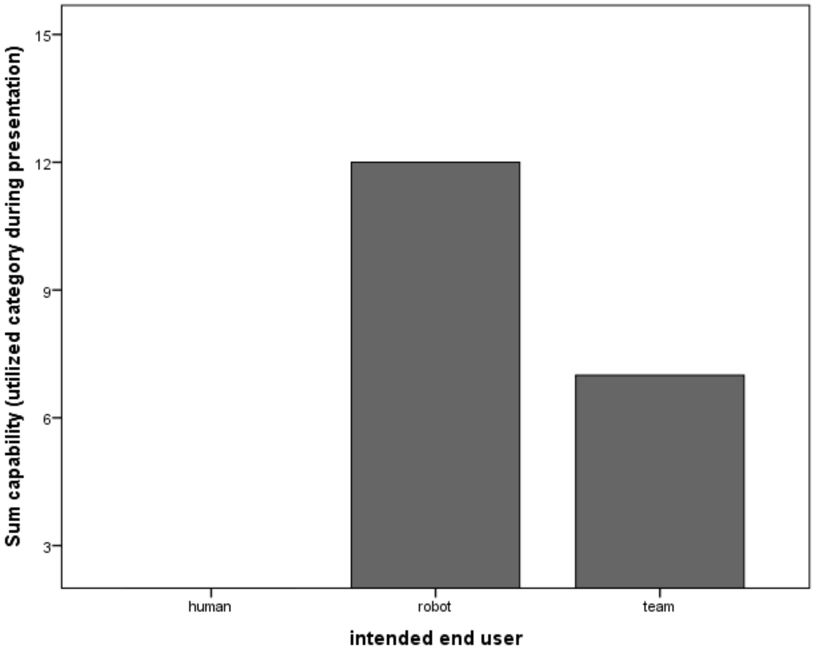

Figure 14. Total number of participants evaluating end user capability during the final tool presentation (*capability* code).

Discussion

We invited engineering students to design a device aiding mundane task completion for one of three end users: a typical person, a robot with unknown technical specifications, and a team consisting of the two. Designers' verbalizations during the design tasks were recorded and explored using Cognitive Discourse Analysis (Tenbrink & Taylor, 2015). We investigated the content of the initial 10-minute brainstorming session and a final tool presentation to investigate approaches participants took to the design process, end user analysis, and design explanation. Given the exploratory nature of our analyses, we advise caution regarding generalizability of our findings.

Process question: how do designers approach the design process?

First, we explored possible differences in how designers approached the design process as a function of the intended end user. User-centered design principles emphasize that a tool's usability is heavily dependent on the end user (Courage & Baxter, 2005). Ignoring end user's requirements can further lead to financial losses due to redesign needs, increased maintenance costs, and possible workplace errors (Marcus, 2002). As such, the end user should play an important role in the design process itself.

First, we investigated different possible starting points to the design task and their impact on design productivity regardless of the intended end user. We defined *productivity* as increased focus on possible tools. Our results demonstrated that between exploring Lego affordances and end user affordances *neither starting point was conducive to focusing more on tool ideas in the long run*. We first analyzed the initial 10 minutes of the design task (brainstorming), and demonstrated that evaluating end user affordances, compared to Lego kit affordances, correlated to more frequent focus on tool ideas. Further analysis of the presentation stage revealed no advantage to either starting point. The initial advantage of focusing on the intended end user was short-lived.

We then explored whether the starting point for the brainstorming session changed depending on the intended end user. *Participants working with robots or teams distributed their focus evenly between assessing Lego affordances and inferring information about their intended end user, while participants working with humans focused almost exclusively on Lego affordances*. In the absence of a valid and applicable mental model of their end user, designers working with robot and team begun by establishing an explicit knowledge base. We based this interpretation on the assumption that if an end user property was already known to be true, it did not need to be explicitly verbalized. By this logic, any user property not explicitly verbalized during the task but critical to the tool design was considered to be known, and not simply omitted by the participant in the design process. Furthermore, our analyses did not explore specific dynamic between the human and robot end users in the team condition. It is possible that the human's involvement in the team allowed designers to shift away from focusing on robot's limitation, instead substituting them with (already known) human capabilities. This would imply that within teams of known and unknown users, lack of knowledge about one user is mitigated by explicit knowledge about the other user.

In an effort to increase cognitive fit (Vessey, 1991) and thus maximize design efficiency and effectiveness, designers ensured that they filled the gaps in their knowledge about the intended end user. In our experimental design, the end user with unknown capabilities was a robot, whereas the end user with known capabilities was a seemingly able-bodied human. However, our paradigm could be expanded by introducing a human end user with unknown capabilities: such as a human with a visible limitation impacting their ability to sort Lego (an arm cast). Introducing a possible limitation to human capabilities would require the designer to reevaluate their mental model of an able-bodied human, possibly leading designers to verbalize more capabilities, limitations, and explanations throughout the design session. Lastly, we investigated whether participants thought differently about the task requirements depending on their end user. We tested whether participants explicitly referred to a specific Lego feature as the most important sorting characteristic, such as shape or size. *The intended end user did not influence whether participants relied on a specific Lego kit feature to aid their sorting.* While working with a less familiar end user required effort to establish a specific knowledge base, it did not detract from focus on task affordances. Participants who worked with robots and teams relied on the Lego feature category to the same extent as participants working with human end users. The task affordances were therefore equally prominent in the brainstorming sessions intended for all end users: designers incorporated some characteristic of Lego blocks, such as shape or size, in devising a feasible sorting strategy.

Similarly, the intended end user did not moderate what kinds of tool designs participant produced. *Participants designing for all end users relied equally on new versus inspired design ideas.* An inspired design idea required inspiration from an already existing tool that served functions similar to those required by the sorting task. Examples included coin sorters or salad spinners, among others. On the other hand, a new design idea did not invoke similar comparisons, implying that the participant was not seeking tool references from experience. We found no differences in how participants approached the tool design task depending on the intended end user. Therefore we concluded that designers did not differ in how they approached the Lego sorting task.

However designers differed significantly in how they approached defining end user capabilities: participants working with humans did not established a clear knowledge base of end user capabilities, implying that such knowledge was readily available and did not need extensive clarification. Mental models of intended end users are crucial for simulating how users could go about completing the task and interacting with the tool; such information in turn serves to identify what kind of assistance the user will need to complete the sorting task, ultimately informing possible tool designs. Applied Cognitive Task Analysis (Militello & Hutton, 1998) serves as one such approach to mental simulation. In our task, designers came in with a preexisting mental model of human capacities that was most likely based on their own capabilities and could be easily generalized to design tasks where the end user appeared to be a typical able-bodied person. Where such mental model fell short of simulating a specific end user, designers acknowledged their limited knowledge and circumvented it by establishing a set of baseline capabilities that then in turn served as a starting point for further elaboration on the end user involvement. Whether the assumptions included in the set of original assumptions were factually correct remains a different question, but the act of establishing the knowledge base alone indicated that designers are aware of the end user role in tool operation, and of the influence of specific end user skills on sorting strategies.

Increasing demand for flexible automation solution led engineers to consider robots as actual tool users and not just tools. As a consequence, further research is needed to evaluate what information about non-human end users is crucial during early stages of tool design. While technical reports describing specific robotic systems should be made available to any engineer in charge of developing tools to be incorporated in flexible automation systems, not all information contained in such reports is immediately pertinent to the initial stages of tool design. Information overload, especially when the designer is not experienced in working with a specific robot, can instead impede performance.

Assumption: how do designers assume and explain information about end users?

Participants designed a tool for one of three possible end users: an average human that the designer could relate to, an unknown robot, and a human-robot team. Manipulating how much our participants knew or could logically infer about their end users allowed us to explore how the existing mental models influenced the design session.

We hypothesized that designers would rely on cognitive heuristics in building mental models of end user capabilities. For instance, designers working with the human end user likely relied on a preexisting mental model of an able-bodied individual built from their own knowledge and experience – paralleling the projective way of knowing described by Nickerson (2001). On the other hand, an absent mental model of the robotic end user required participants to acquire knowledge on end user capabilities through other means, such as visual inspection. Participants could identify a possible anchor – a starting point to building plausible estimates (Epley & Gilovich, 2006) – and infer further end user capabilities on its basis. We interpreted our data with an assumption that verbalized assumptions, limitations, or actions required for sorting indicated that the designer was filling in gaps in their mental representation.

Different intended end users prompted designers to treat them differently. When designing for a robot, participants emphasized the role that the specific robot needed to play in operating the tool or sorting the Lego kit by invoking the *action* category.. Designers working with human end users instead tended to explain their sorting strategies and tool ideas as nondescript procedural expressions by invoking the *strategy* category, failing to directly address the human end user that was sitting across the table. For instance, if the end user needed to put a Lego block in a specific compartment, an action description would explicitly state that involvement ("Baxter then puts this piece in this container"). On the other hand, the same situation described as a strategy would not implicate a specific end user, and instead be expressed either as "this Lego piece gets put in this container" or "you would put this piece in this container."

When we ignored the explicitly stated end user involvement and instead consider action and strategy descriptions to be qualitatively identical, any differences in how users relate to different end users disappeared. *All end users were equally important to the design process, but their role in the sorting process was approached differently.* A comparison of action descriptions and strategy descriptions within the team condition – that is, how participants differentiated roles within a robot-human team – revealed a similar pattern. When the difference between action and strategy was removed, both end users in the team condition received the same amount of focus from the designers. These results suggest that designers working with humans have no difficulty generalizing their end user mental model to apply to the task at hand. They also have no problem generalizing their final solution to be used by all possible human end users, not only the confederate sitting across the table. Such flexibility emphasizes how strong the mental model of a human end user can be. On the other hand, participants working with robots could not rely on such a flexible framework – their end users were directly implicated in sorting both during the brainstorming stage and presentation stage, indicating lack of flexibility and generalizability of their established knowledge base to other possible hypothetical robots. Mental model's flexibility and generalizability should be explored further by introducing a wider variety of robotic end-users. Our study explored the design process that includes only one robot, and as such cannot inform any possible mental model transfer and adjustment between different robots.

A closer analysis of how designers thought about their end users revealed that *designers elaborated on end user capabilities and limitations more explicitly for the robotic or team end user as compared to human,* further confirming our claim that designers must first establish a knowledge base on end users that fail to be accounted for by the designers' existing mental models. Additional analyses are needed to explore the dynamic between human and robot end users within the team condition. Such knowledge base was also subject to critical evaluation: *designers working with robots and teams were more frequently critical of their end user abilities, as compared to designers working with humans.* However this difference was not driven by a sampling error that placed all critical designers in robot or team conditions. *Designers working* *with humans were more critical of their own design ideas, as compared to robot or team designers.* The criticism was targeted towards their ideas instead of the end user as their mental model of end user is pretty strong and includes both end user strengths and weaknesses. If a limitation is known, then there is no need for an outright questioning of end user's abilities. Contrary to that, stating a limitation of an unknown end user is as important as stating a capability, as it contributes to building an explicit mental model of a robot end user.

Shared knowledge: how do designers explain their designs?

We demonstrated that designers lack knowledge of robotic end user capabilities, and must first establish a knowledge base on their end user before exploring possible problem solutions. Specific mental models and cognitive heuristics assist the designer in designing a tool, and as such play an important role in explaining the final solution: the designers construct the end user's mental model during the brainstorming session, and restate their assumptions during the presentation portion. Successful communication requires designers to ensure that their audience possesses all the knowledge necessary to understand the tool solution. Designers appeal to the common ground (Clark & Carlson, 1981; Clark & Marshall, 1981) with the audience, which effectively is another mental model – this time, of shared knowledge. During conversation, common ground can be inferred through interaction between the speaker and the addressee (Isaac & Clark, 1987). During presentations, designers have to assume the common ground and therefore adjust their explanations accordingly to ensure that the supplied expertise is accessible to a range of possible listeners (Schober, 1993).

We investigated how our participants communicated their tool solutions to a larger audience, and concluded that designers assume almost no common ground with their audience when discussing robotic end user capabilities as opposed to human end user capabilities. In other words, they assumed their audience knew little about the robot end user. *Participants took significantly longer to explain their tool idea designed for a robotic and team end users, as opposed to human end users*. We asked participants to present to a hypothetical audience while still seated in front of their intended end user. Human confederate's presence could warrant a perspective shift where participant explains the design to the confederate instead of an actual hypothetical audience during human and team conditions. Schober (1993) argued that people tend to be more verbose during monologues as compared to conversations when communicating knowledge. If this was the case in our study, participants presenting for humans and teams would produce presentations of roughly the same length, as a human confederate was present in both conditions. However our participants still needed significantly more time to explain their design within the team condition as compared to human condition. We therefore concluded that confederate's presence did not seriously affect the presentation stage. We compared talking speed between three end user groups to account for a possible sampling bias, with participants speaking more slowly being assigned to the robot condition. We found no evidence of such bias and therefore concluded that the increased time required to explain their tool design was moderated by the need to explain why certain design choices were made with regard to a specific end user. In fact, *participants working with robots and teams were more explicit about naming their end user's direct involvement in the sorting process,* invoking the *action* category and emphasizing the inflexibility of the robot mental model.

Participants designing for robots and teams explicitly discussed end user capabilities as compared to participants working with humans, but neither end user group focused on limitations during the presentation stage. It appears that limitations, while important for establishing a knowledge base in early stages of design, were not as relevant to presenting a tool idea. Our findings were in line with Grice's (1975, 1978) maxims of quantity and relevance: participants cooperated with their hypothetical addresses, and as such their presentations of end user capabilities were immediately relevant to the task at hand, and presented in an informative fashion but without overloading the addressee with unnecessary details. Limitations are crucial for design as they narrow down the possible end user capabilities, which in turn contributes to establishing a knowledge base during design. On the other hand, limitations are not immediately pertinent to the explanation of the sorting tool – because a tool will most likely play to the end user strengths and capabilities, and therefore explaining end user limitations would lead to overloading the audience with irrelevant details and detracting attention from the actual tool solution.

Conclusion

We demonstrated that the end user played an important role in how designers approach designing tools. While the end user did not affect how participants conceptualized task affordances or approached their final solution, we demonstrated differences in how participants approached the role and the importance of the end user in their final solution. Designers working with humans exhibited a flexible mental model of their end user that they assumed to be common ground with a possible audience. However designers working with robots and teams needed to establish a knowledge base about their end user, and communicate it explicitly – that is, verbalize - to a possible audience to establish common ground.

Our data demonstrated that during early stages of tool design, participants relied heavily on their preexisting mental models in assessing end user capabilities. Human factors design provides extensive details on physical and cognitive skills of human end users, yet our designers who worked with human end users did not indicate any reliance on existing human factors research during the design session. As we only explored the initial brainstorming (first 10 minutes of the design process), it is possible that the preference for mental models over hard data was driven by time limitations or the general experimental setting (participants were not required to build a functional, testable prototype). Such circumventing was not possible when designers worked with a robot or team – as no information about the end user was at all implicit and could be referenced as easily as a mental model of a human end user. Unfortunately, we lack knowledge of what kinds of information are pertinent when a designer is working with an unknown robotic end user. We therefore pose the need for what we term *robot factors design*: a variation on human factors design evaluating what depth of information about the robotic end user is necessary and sufficient during early stages of tool design. The challenge in robot factors design, as opposed to human factors design, is that the variety of flexible automation systems available on the market is too great to propose a fixed set of robot characteristics required to effectively address task-specific tool design. Therefore we propose that it is not the importance of *specific characteristics* of flexible automation systems that must be investigated through robot factors design in order to aid working with such end users, but what *types* and what *depth* of information presented to the designer at early stages of tool design contributes to design productivity. Robot factors design would be an intermediate step between designers working with completely unknown end users and designers working with informational overload about their end user, but would provide information on the most effective starting points for creating end user mental models.

Table 1. Coding scheme: five main categories with respective sub-categories and definitions used in our analysis.

Category	Sub-category	Definition	Example
End user		utterances related to preexisting or standalone (i.e., not required by the context) capabilities	
	ability	explicitly stated (verbalized) known capabilities of the end user	“the one thing that um Erica has that Baxter hasn't that doesn't have is um the ability to touch and feel objects and if you crush”
	assumption	explicitly stated (verbalized) possible capabilities of the end user	“first of all I am going to assume that Baxter is going to be able to recognize each of these pieces himself”
	comparison	comparing robot's capability to human's	“this way like a wrist”
	details	elaboration on anything in this category	“there are cameras at the end of Baxter’s hands”
	explanation	elaborating why the participant claims a particular assumption, limitation, or ability	“because of the way his arm is stretched with all the joints”
	limitation	explicitly stated (verbalized) lack of capabilities of the end user	“looking at Baxter’s hands they appear to not have that much fine motor control”
	uncertainty	an uncertain stance on end user's ability	“does Baxter have like those capabilities to see color and that sort of thing?”
Lego kit		utterances related to evaluating the contents of the NXT kit	
	color		“Now as it happens the pins ones I suppose that they're a darker gray and these are the lighter gray”

	container		“it’s not like immediately two different containers”
	general	other Lego affordances	“basically all of one set are the same”
	material		“pieces have rubber, tires have rubber as opposed to the other pieces”
	number		“I have one, two, three, four small, two large and three- about three sensors”
	organization		“more of the larger pieces are like in certain compartments in here”
	reference image		“in terms of the reference picture”
	shape		“L shaped things”
	size		“there’s some longer ones”
	type		“the other one looks like there’s some like smaller gear type looking things with some other connecting pieces”
	weight		“because the mass of all of these is kind of similar”
Design		utterances pertaining to what the final design could, should, or will look like, including a list of tasks to be performed by end users	
	action	explicitly stating (verbalizing) an action that a specific end user needs to perform in order to operate the tool and/or contribute to the sorting process	“then picks each thing one by one”

	capability	explicitly stating (verbalizing) what kind of capability is crucial for successful tool usage or participation in sorting	“or if he could recognize all of them he could scoop them up”
	comparison	comparing the tool idea to an existing tool (for example, coin sorter)	“cause it’s just like that just like how your own fingers are rounded”
	current	reiterating the current stage of the sorting process	“so everything’s on the table”
	feature	addressing a specific Lego affordance crucial for successful sorting	“if everything was magnetized that would be easier”
	limitation	limitation of the tool	“then in this case something else could fall over here and then just and mess it up even more”
	operation	how tool needs to be operated - without specifically addressing the role of the end user	“that could just be clamped on anyway you want”
	purpose	reason for certain adjustments in the process	“he can pick up tinier things easier with these tools”
	software	emphasizing the need to reprogram the robot	“it would some sort of like matching like software that he could see”
	strategy	stating an action needed to aid the sorting process without specifically addressing the end user	“then if it does go through and it can like get into a hole”
	tool	reiterating design idea	“so you would need something that you could attach here”
Process, Evaluation, Problem		utterances where the designer is evaluating her thinking, actions, or identifies challenges within the design	

accept idea	accepting an idea or aspect of an idea	“that might be a good metric”
accept reason	reason for accepting idea or aspect of an idea	“because it’s where the other pieces are”
consider alternative	admitting possible alternative solution paths	“there are other ways to sort things”
demonstration	explaining how the tool works, usually accompanied by a physical demonstration	“what I am doing right now is trying it out with my own hands seeing the grip that I need”
dismiss issue	acknowledging an issue but dismissing it as unimportant	“but that doesn’t matter”
evaluate idea	evaluating limitations and advantages of an idea	“that would make my life easier”
evaluate kit	evaluating whether the Lego kit has properties compliant with the tool idea	“but most of the pieces that belong in here seem to be very easy to pick out”
evaluate user	evaluating whether the end user can operate the tool	“I’m not going to worry about you because you have a lot of mobility”
experience	referring to designer's specific experience	“I always try to, I just kinda dump them and then sort them in different places and then put them in the kits”
identify idea	acknowledging a possible solution path	“maybe using that somehow and tension?”
improve capability	directly addressing a limitation of end user	“I feel like it would probably be best to use both of his hands for two different tools”
limited knowledge	acknowledging limited knowledge of user, kit, or problem	“I’m not exactly sure how his hand works”

	naming	acknowledging lack of appropriate vocabulary	“double bushing is what I think they’re called”
	problem	identifying a problem or difficulty	“but how would he control based off of just those two motions?”
	question idea	deliberating whether an idea is plausible	“would that help in any way?”
	question user	deliberating whether an end user can handle operating the tool	“would he be able to pick this up?”
	reject idea	reject an idea or aspect of an idea	“I can’t put something sticky on the end”
	reject reason	reason for rejecting idea or aspect of an idea	“’cause it’s not going to pick it up”
	table issue	identify an issue but deciding to not deal with it at the time	“we’ll fix that in a little bit”
	thinking	narrating current actions (such as sketching or moving)	“I’m just gonna draw the arm”
Reiterate and clarify task		any utterance pertaining to clarification, including questions directed at the researcher	

References

Argote, L., & Ingram, P. (2000). Knowledge transfer: a basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes, 82*(1), 150-169.

Armstrong, J.S., Denniston, W.B., & Gordon, M.M. (1975). The use of the decomposition principle in making judgments. *Organizational Behavior and Human Performance, 14*(2), 257-263.

Banks, V.A., Stanton, N.A., & Harvey, C. (2014). What the drivers do and do not tell you: Using verbal protocol analysis to investigate driver behavior in emergency situations. *Ergonomics, 57*(3), 332-342.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour and Information Technology, 13,* 132-145.

Blackburn, J.M. (1936). *The acquisition of skill: An analysis of learning curves.* London: Industrial Health Research Board.

Carroll, J. M., & Olson, J. (Eds.). (1987). *Mental models in human-computer interaction: Research issues about what the user of software knows.* Washington, DC: National Academy Press.

Cheong, H., Hallihan, G.M., & Shu, L.H. (2014). Design problem solving with biological analogies: A verbal protocol study. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 1,* 27-47.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long and A. Baddeley (Eds.), *Attention and performance IX* (313-331). Hillsdale, NJ: Erlbaum.

Clark, H. H., and Haviland, S. E. (1974). Psychological processes in linguistic explanation. In D. Cohen (ed.), *Explaining Linguistic Phenomena* (91–124). Washington: Hemisphere Publication Corporation.

Clark, H. H., and Haviland, S. E. (1977). Comprehension and the Given–New contract. In R. O. Freedle (Ed.), *Discourse Production and Comprehension* (1–40). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber and I. A. Sag (Eds.), *Elements of discourse understanding* (10-63). Cambridge, England: Cambridge University Press.

Coeckelbergh, M. (2011). *Can we trust robots? Ethics and Information Technology, 14*(1), 53-60.

Courage, C., & Baxter, K. (2005). *Understanding your users. A practical guide to user requirements, tools, and techniques*. Elsevier: San Francisco, CA.

Dietvors, B.J., Simmons, J.P., & Massey, C. (2014). Algorithm aversion: People erroneously avoid algorithms after seeing them Err. *Journal of Experimental Psychology: General, 144*(1), 114-126.

Durfee, D.H., Boerkoel Jr., J.C., Sleight, J. (2014). Using hybrid scheduling for the semi-autonomous formation of expert teams. *Future Generation Computer Systems, 31,* 200–212.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic. *Psychological Science, 17*(4), 311-318.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis; Verbal reports as data (revised edition). Cambridge, MA: Bradford books/MIT Press.

Erlandsson, M. & Jansson, A. (2013). Verbal reports and domain specific knowledge: a comparison between collegial and retrospective verbalisation. *Cognition, Technology & Work, 15*, 239-254.

Forlizzi, J., & Batttarbee, K. (2004). Understanding experience in interactive systems. *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques.* New York, NY: USA.

Forrester, J.W. (1971). Counterintuitive Behavior of Social Systems. *Technology Review, 73*(3), 52-68.

Gombolay, M.C., Gutierrez, R.A., Clarke, S.G., Sturla, G.F., & Shah, J.A. (2015). Decision making authority, team efficiency and human worker satisfaction in mixed human-robot teams. *Autonomous Robots, 39*(3), 293-312.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics, Vol. 3: Speech Acts* (113–28). New York: Seminar Press.

Grice, H. P. (1978). *Some further notes on logic and conversation*. In P. Cole (Ed.), Syntax and Semantics 9: Pragmatics: 113–27. New York: Academic Press

Halliday, M. A. K. (1967). *Notes on transitivity and theme in English. Part 2*. Journal of Linguistics, 3, 199–244.

Hancock, P.A., Billings, D.R., Schaefer, K.E., Chen, J.Y.C., de Visser, E.J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*(5), 517-527.

Haring, K.S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. *Proceedings of the World Congress on Engineering and Computer Science*. San Francisco, CA: USA.

Hassenzahl, M., & Tractinsky, N. (2006). User experience – a research agenda. *Behaviour and Information Technology, 25*(2), 91-97.

Haviland, S. E., and Clark, H. H. (1974). *What's new? Acquiring new information as a process in comprehension.* Journal of Verbal Learning and Verbal Behavior, 13, 512–21.

Hinds, P.J. (1999). The curse of expertise: the effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied, 5*(2), 205-221.

Hoffman, K.A., Aitken, L.M., & Duffield, C. (2009). A comparison of novice and expert nurses' cue collection during clinical-decision making: Verbal protocol analysis. *International Journal of Nursing Studies, 46*(10), 1335-1344.

Johnson, J. (2010). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules.* Burlington, MA: Morgan Kaufman Publishers.

Kankuzi, B., & Sajaniemi, J. (2016). A mental model perspective for tool development and paradigm shift in spreadsheets. *International Journal of Human-Computer Studies, 86*, 149-163.

Kozbelt, A., Dexter, S., Dolese, M., Meredith, D., & Ostrofsky, J. (2014). Regressive imagery in creative problem-solving: Comparing verbal protocols of expert and novice visual artists and computer programmers. *The Journal of Creative Behavior, 49*(4), 263-278.

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd Ed.). London: Sage.

Kuusela, H., & Pallab, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *The American Journal of Psychology, 113*(3), 387- 404.

Lee, S.W., Park, J., Kim, A.R., & Seong, P.H. (2012). Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis. *Annals of Nuclear Energy, 43,I 167-175.*

Maguire, M. (2001). Methods to support human-centered design. *International Journal of Human-Computer studies, 55*, 587-634.

Marcus, A. (2002). Return on Investment for Usable UI Design. *User Experience Magazine*, 1(3), 25–31.

Militello, L.G., & Hutton, R.J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.

Newell, A., and Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nickerson, R.S. (2001). The projective way of knowing: A useful heuristic that sometimes misleads. *Current Directions in Psychological Science*, *10*(5), 168-172.

Proctor, R.W., & Vu, K.P.L. (2016). Principles for designing interfaces compatible with human information processing. *International Journal of Human Computer Interaction, 32*(1), 2-22.

Reimann, P., & Chi, M. T. H. (1989). Expertise in complex problem solving. In K. J.Gilhooly (Ed.), *Human and machine problem solving* (pp. 161-192). New York, New York: Perseus Books Group.

Robie, C., Brown, D.J., & Beaty, J.C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Managerial Psychology, 18*, 156-165.

Robinette, P., Li, W., Allen, R., Howard., A., & Wagner, A. (2016, March). *Overtrust of robots in emergency situation scenarios*. Paper presented at the 11th ACM/IEEE International Conference on Human-Robot Interaction, New Zealand.

Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349-363.

Schober, M.F. (1993). Spatial perspective-taking in conversation. *Cognition,* 47, 1-24.

Senders, J.W. (2006). On the complexity of medical devices and systems. *Quality and Safety in Health Care, 15*, i41-i43.

Someren, W. M., & Barnard, F. Y., & Sandberg, A.C. J. (1994). The think aloud method. A practical guide to modeling cognitive process. London: Harcourt Brace & Company.

Tenbrink, T. (2014). Cognitive discourse analysis: Accessing cognitive representations and processes through language data. *Language and Cognition, 7*(1), 98-137.

Tenbrink, T., & Taylor, H.A. (2015). Conceptual transformation and cognitive processes in origami paper folding. *Journal of Problem Solving, 8*, 2-22.

Verhoef, L.W.M. (2007). Why designers can't understand their users: Developing a systematic approach using cognitive psychology (Doctoral dissertation). Retrieved from Leiden University Repository. (ISBN 9789080997516)

Vessey, I. (1991). Cognitive fit: a theory-based analysis of the graphs versus tables literature. *Decision Sciences, 22*, 219-241.

Vessey, I., & Galletta, D. (2001). Cognitive fit: an empirical study of information acquisition. *Information Systems Research*, *2*(1), 63-84.

Ward, P., Suss, J., Eccles, D.W., Williams, A.M., & Harris, K.R. (2011). Skill-based differences in option generation in a complex task: A verbal protocol analysis. *Cognitive Processing, 12*(3), 289-300.

Wickens, C.D., Lee, J., Liu, Y., & Brecker, S.G. (2004). *An introduction to human factors engineering* (2nd edition). Boston, MA: Pearson.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.

Yanco, H.A., & Drury, J. (2004). Classifying human-robot interaction: An updated taxonomy. *IEEE International Conference on Systems, Man and Cybernetics*, 3, 2841-2846.

Footnotes

---

[1] In order to easily distinguish between robots-as-tools and robots-as-tool-users in this context, we will rely on the following vocabulary distinction: the term “tool user” (or simply “user”) refers to either a robot or a human using a particular device. However when a robot is used by a human as a tool, we refer to the human as “operator.” Therefore tools are used by humans or robots, but robots are operated by humans.