Wizard of Awwws: Exploring Psychological Impact on the Researchers in Social HRI Experiments

Daniel J. Rea, Denise Geiskkovitch, James E. Young
University of Manitoba
{daniel.rea, young}@cs.umanitoba.ca, umgeiskk@myumanitoba.ca

ABSTRACT
In social Human-Robot Interaction (sHRI) people have studied social interactions with awkward, confrontational, or unsettling robots. In order to create these situations, researchers often secretly control the robot (the "Wizard of Oz", WoZ, technique), use confederates (researchers pretending to be participants), or the researchers themselves create the desired social condition. While these studies may be antagonistic, they are designed to be ethical; when conducting a study, IRB (Institutional Review Board) processes are in place to assess the study design for potential risk to participants, and to ultimately protect the public. However, these processes do not generally involve assessment of impact on the researchers conducting the study. In our own work, we have noted how researcher "wizards" in social HRI experiments, particularly those which place participants in awkward or confrontational situations, can themselves be negatively impacted from the experience when their experiment protocol has them antagonize, deceive, or argue with participants. In this paper, we explore how experimental design can impact the wellbeing of the researchers, particularly for wizards in social HRI experiments. By building a psychological grounding for the impact on people who do socially stressful actions, we evaluate the potential for researcher social stress in recent sHRI studies. Our summary and discussion of this survey results in recommendations for future HRI research to reduce the burden on wizards in their own experiments.

CCS Concepts
• Human-centered computing→Human Computer Interaction (HCI)→HCI design and evaluation methods→User studies

Keywords
Human-robot interaction; social stress; researcher stress; experiment design; ethics

1. INTRODUCTION
Studies involving human participants are a core component of research in Human-Robot Interaction (HRI), particularly work that focuses on social interaction between humans and robots – social HRI (sHRI). Researchers involved in sHRI experiments are often themselves part of the social interaction, for example, by interacting with participants and the robot, sometimes as the authority figure controlling the experiment. Researchers can also be confederates ("fake" participants interacting with real participants), or "wizards," secretly controlling a robot to interact socially (the "Wizard of Oz", (WoZ) technique [29]). These techniques are common-place as they enable the investigation of specific social situations and interactions between people and ostensibly intelligent robots. To protect participants in these social experiments, procedures are in place to assess the ethics of a study design, but less consideration has been given to how these social interactions may negatively impact the researchers.

Figure 1. A wizard-controlled robot pressures a participant in an experiment (image from [13]**). Based on psychology literature, such behavior may be stressful to the experimenters as well.**

The use of formal studies involving participants to investigate social situations is commonplace in broader Human-Computer Interaction (HCI), such as in affective computing and virtual agents. sHRI is unique in this area given robots' physical embodiments, which helps create a particularly strong sense of agency: this raises the realism and impact of social and emotional responses that people may have when interacting with them [51]. Further, sHRI has a history of testing the boundaries of confrontational or socially awkward situations involving robots, echoing classic psychology work (e.g., the Milgram [38] or Stanford Prison [21] experiments). For example, researchers have had robots pressure participants into doing uncomfortable tasks [1], continuing a task even after participants ask to stop [13], or to do morally-questionable tasks [2,3]. We believe that robots have physical and social realism [32,51], thus socially uncomfortable interactions with them are psychologically similar to analogous interactions with people; this may create a negative social and emotional impact on researchers, thus highlighting the importance of considering how such sHRI experiments may impact the researchers who conduct them.

In this paper, we propose that sHRI experiments have the potential to negatively impact the researchers conducting them; it is reasonable to expect researchers to empathize with a participant undergoing stress, relate to the social experience in the experiment, or feel stress themselves from creating a stressful situation for others. Research in Psychology details how socially uncomfortable

HRI '17 Companion, March 06-09, 2017, Vienna, Austria
 ISBN 978-1-4503-4885-0/17/03..$15.00
DOI: http://dx.doi.org/10.1145/3029798.3034782

situations in general can cause stress [4,5,23,30], and how common experimental design components such as deception (such as being a robot wizard or confederate) can amplify this [18,21,22]. Indeed, in our own work, we have noted that such social experiments sometimes have a negative impact on a wizard's wellbeing, particularly when the research protocol has the wizard being antagonistic. Further, researcher stress may negatively impact experiment results by reducing performance and ability to maintain any required deception [18].

This paper is a call for awareness, research, and inquiry into socially induced stress for researchers in sHRI experiment design. We first outline a basic psychological framing for potential social stress for researchers conducting studies, and then analyze a number of recent sHRI studies for aspects which may induce such stress. We finish with initial recommendations for how future research can take care to reduce – or at least be aware of – potential researcher stress in sHRI experiments.

2. RESEARCHER STRESS

There is a body of research exploring the general problem of HRI evaluation [42,51]. Work that considers potential impacts of research study design on people has generally focused on the participants (e.g., [36]), such as the deception of participants [43]. Some even consider the ethics of how robots are treated [24]. However, we are only aware of one work that has called for a consideration of the impact of study design on researcher wellbeing, which specifically called for clearly informing researchers of the potential for stress when they will be placing participants in uncomfortable social situations [13]. As such, we see a hole here that should be filled. We call for increased research on how study designs may negatively impact researchers. Our work in this paper begins to address this by highlighting potential issues, and analyzing recent sHRI work, examining how experiment design impacts researcher stress.

First, we explore the evidence for how and why researchers may be affected by their experiments. We then explore general stress in social interaction from the lens of a researcher conducting a sHRI experiment and expand on work in Psychology specifically pertaining to researcher wellbeing. Finally, we expand on the case of Wizard of Oz studies, and detail why this may be particularly stressful to researchers.

2.1 Researchers Potentially Impacted by Participation in Studies

The Stanford prison experiment [21] is a seminal study contributing to modern experimental ethics on many fronts. While it is commonly known for the impact it had on research participants, it also clearly documented and demonstrated that researchers themselves could become psychologically affected by being involved in the experiment. The well-known goal of this work was to explore psychological power dynamics between participants acting as either prison guards or prisoners. However, the principle investigator, who was involved by acting as the prison superintendent, became absorbed into the role and was impacted by the social situation of the experiment. In this case, as a result the researcher ultimately allowed and enabled clearly unacceptable psychological abuse between participants, a fact that other researchers involved have noted caused them a lot of stress [52]. Thus, even experienced researchers who themselves designed an experiment, can be psychologically affected by those same experiments.

An important result of the Stanford Prison Experiment is that both the researchers and participants did not stop the experiment – even though they could choose to do so – despite how stressful it was for those involved (though a few participants quit partway, many stayed). The principle investigator hesitated to stop the experiment even after external observers highlighted the unacceptable abuses, though the external observers convinced him in the end [21,52]. In this case, it clearly demonstrates that researchers – as well as participants – can be prone to *situational attribution of behavior*, people will often act in ways that suit a situation, even if it may go against their natural disposition [21]. While we are careful to note that the majority of sHRI experiments will not result in this strong case of directly harming participants, the principle holds in the more general case. That is, we can expect researchers involved in social interactions – even when they know they are acting – to have a tendency to behave and feel as if the situation is real. If the researcher is partaking in negative social interactions, then this may have negative impact on the researcher's mental wellbeing.

Follow up work to the Stanford Prison Experiments is sparse, particularly with respect to impact of study design on researchers. Notable examples include the consideration of ethics in medical simulations involving death (virtual deception) [14] and that the interplay of power and status may affect how researchers and participants treat each other [16]. These projects, when talking about harm, are primarily concerned with participant health; we focus on considering this potential risk to researchers in the context of sHRI studies.

While there have been no studies, to our knowledge, in the vein of the Stanford Prison Experiment in sHRI, there have been several studies that have pushed the boundaries of the field by studying socially uncomfortable situations. Echoing the Stanford Prison Experiment [21] and Milgram's obedience study [38], some sHRI researchers have, for example, studied obedience to robots when compared to humans [13], or have used their authority as a researcher and told participants to destroy robots [2,3], or perform embarrassing medical procedures [1]. We propose that, like the Stanford Prison Experiment, these researchers may have felt social stress arising from these confrontational situations; we explore this point by looking at the psychological basis for where social stress could arise, and then apply that knowledge to sHRI studies that examine uncomfortable social interactions with robots.

2.2 General Stress in Social Interaction

We explain here how the social situations that researchers can partake in sHRI work can reasonably be expected to be stressful. In psychology, research has investigated how stress can arise in social situations [4,23], how people cope with social stress [30], and the effects stress can have on people [11,20,41]. Such research provides insights into how stress could affect researchers during and after sHRI experiments.

In social interactions, people seek frequent, positive interactions to form a sense of belonging; frequent negative reactions can result in anxiety and stress [4]. While this behavior is typical when interacting with close friends, people can also quickly and easily bond with others they have only just met [18]; thus it is possible to feel connected to and be affected by repeated interactions with someone you've just met, such as a researcher interacting with a participant. If those repeated interactions are negative, unexpected, or uncontrollable, such as discrimination, physical stress, or arbitrary and confrontational, there may be a potential for stress to the researcher [12].

Researchers, particularly confederates and wizards in WoZ experiments, are placed in a situation similar to participants: they may feel a loss of control and helplessness as they are following a scripted act that enforces a potentially confrontational or antagonizing protocol (e.g., [1,3,13]). These feelings of helplessness and lack of control can lead to stress and anxiety [12,23].

2.3 Stress in sHRI Studies

Cognitive demand on researchers is common in many experiments, but can combine with social stress to produce negative effects [12]. For example, deception, such as misleading participants in order to create desired social situations to study, or being a confederate or wizard, is common in sHRI. However, employing deception can raise cognitive stress: deceit takes more cognitive processing to maintain, and the risk of failure of the deception can result in additional mental and physical stress [18,41]. Researchers involved in sHRI evaluations often are observing a broad range of interactions, taking a holistic view of the interaction, including personal instincts and feelings, body language and word choices, and the overall evolution of a social interaction [51]. Keeping this number of parameters in check, especially while maintaining deception [18], can increase overall cognitively load [12].

Once stressed, a person applies coping strategies to manage the stress [30]. While how one copes in a situation is heavily dependent on their emotional intelligence and empathy [17,46], there are common broad-brush strategies that are relevant to researchers in sHRI. In formal experiments that follow scripts and procedures, researchers have little leeway in how they can alter their behavior. As such, they are likely to employ avoidance coping (distracting oneself from how one feels), or emotional coping (focusing on the negative feelings to attempt to diffuse them); neither of these are effective for maintaining mental health [11,15,20]. Task-focused coping, where a person actively tries to change the stressful circumstances, is much healthier [30]. While scripted experiments limit how this can be employed immediately, research design can create positive interactions quickly, such as having a researcher meeting and talking with participants immediately after an experiment, [4].

2.4 Potential Consequences of Researcher Stress

Researchers who create stressful social situations can also experience that stress themselves [21]. It has also been shown that stress arising in laboratories can have similar psychological and physiological effects as analog situations in natural environments [12]. This has broad health implications, as stress can impact physical health [4,20,41], interpersonal relationships [4,20], and cognitive ability [41].

Stress can affect the performance of researchers in their experiments, and thus the outcomes and reliability of research results. Research has demonstrated how emotional suppression (e.g., a common coping strategy) can reduce someone's ability to communicate [11,12] by increasing cognitive demand [41], potentially impacting study quality. Evidence also suggests stress can make people socially insensitive [12], an important consideration of impact for observation-based sHRI studies. Further, while we saw above how including deception in a study can increase stress (see above), stress can further impact the quality of deception. People who believe lying is morally impermissible tend to have lower deception abilities when under stress [18]. When stressed, deception can become more obvious to observers, and believability is lowered [18].

2.5 Stress in Wizard-of-Oz Studies

Wizarding in particular has some stressors that may increase the stress of the researcher. In sHRI, some experimental designs include an initial phase to build rapport and empathy between the participant and the robot before starting the experiment (e.g., [33,44,47]), and negative interactions with in-group relations can increase stress for the researcher [4]. Wizards may feel isolated as they operate the robot remotely in another room, which can exacerbate feelings of stress [23]. The remote aspect may again cause stress due to embarrassment from eavesdropping [34]. Additionally, controlling a robot can be a highly cognitively demanding task [19,48], further increasing the mental load for wizards. Together, these combine with the above stressors to make wizarding potentially stressful in socially uncomfortable experiments.

In the below sections, we group papers by rough themes that highlight particular ways that the experiment may be stressful to researchers. Each paper is given its own section, with the paper title (and citation) as the section name.

3. ANALYSIS: POTENTIAL STRESS IN RECENT HRI STUDIES

In this section we analyze a selection of recent sHRI works from the perspective of considering researcher wellbeing from involvement in a study, using our understanding of human stress outlined in the prior section. The purpose of this analysis is to explore, in the context of real work, aspects of study design that may have the potential for impacting researcher wellbeing. Conversely, we aim to identify study designs which may protect researchers. We surveyed papers in the ACM/IEEE International Conference on Human-Robot Interaction, the ACM International Conference of Human-Agent Interaction, the Journal of Social Robotics, and the Journal of Human-Robot Interaction.

This is not an exhaustive survey, as our goal is to find exemplar works for analysis purposes. Further, we emphasize that *we do not aim to criticize these works*, either from a scientific or moral standpoint. Our goal is rather to use these as examples to aid us in exploring the issue of researcher wellbeing when conducting sHRI studies.

3.1 Persuasion, Obedience, and Social Pressure

One group of work studies how people react in the face of authority, whether it is a robot giving the orders, or a human researcher ordering a participant to perform some socially unacceptable act to a robot. The core social stress in these works stems from participants being ordered or continuously pressured to perform some task, even if they protest. This may create a negative social interaction, and thus stress, for the researcher (or WoZ robot) [4].

3.1.1 Would You Do as a Robot Commands? [13]

This project investigated how people react when repeatedly pressured by a robot or a human to complete a tedious task, even after repeatedly asking to quit (Figure 1). Researchers were involved in three relevant capacities: as a human pressuring participants in one of the conditions, the human introducing and orchestrating the experiment, and placing the participant in the situation, and the human acting as a wizard remotely controlling the robot, and directly pressuring people.

The primary concern for researcher stress in this experiment is social guilt arising from placing the participant uncomfortable situation. A large part of this arises from the confrontational social

situation that the researchers partook in for 27 participants. When a participant indicated that they wanted to quit the experiment, the researcher (directly, in the human condition, or as a wizard, in the robot condition) followed a strict confrontational protocol to insist the participant continue. The paper reports that this resulted in participants being visibly upset, arguing, whining, displaying exasperation, calling the robot names, and other behaviors that indicated their stress. Not only do we expect general researcher stress from the negative social interactions [4], this may be compounded by a potential feeling of helplessness given the need to act according to a strict protocol script [30], and the physical isolation for the wizard operator.

At the end of this experiment, the protocol immediately had a friendly reconciliation and debriefing between the researcher and participant. This enabled the researchers to apologize, have a positive interaction, make sure the participant was okay, and end the deception as soon as possible; such active and positive social interactions can be expected to reduce stress [4,30].

3.1.2 The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots [1]

Researchers in this work had various medical robots that could convince participants to perform increasingly embarrassing medical procedures (Figure 2). Researchers introduced the study and debriefed participants in person, but were otherwise in another room, controlling the robot as a wizard. Potential stress could come from empathy when the 44 participants seemed uncomfortable with the procedure, or real embarrassment from seeing participants undress, typically a private matter [34]. There is also a potential for future discomfort if they meet participants later, e.g., on campus. Participants were debriefed, but no specific details on how the researcher may have tried to create positive social interactions were described.

3.1.3 To Kill a Mockingbird Robot [3]

This project studied people's disposition to be violent towards a robot based on the robot's intelligence. Researchers were present in person to guide the interaction, and pressured people at the end to smash the robot with a hammer until the robot stopped moving.

The primary concern for stress comes from social guilt from making participants uncomfortable by pressuring them to commit violence on the lab equipment. If any of the 16 participants hesitated or refused, or did not do enough damage, the researcher followed a protocol to pressure them into continuing. Such confrontation is a negative social interaction which could cause stress [4], potentially compounded by the inability to break from protocol if the participant was uncomfortable [30]. After debriefing participants, researchers inquired if the participants felt any negative effects, creating a helpful and positive interaction which could reduce researcher guilt or stress levels.

3.1.4 Summary of Stressors

In these works, the common stressors to the researchers were: 1) guilt from forcing participants to comply with uncomfortable orders, and 2) stress from witnessing or being unable to respond to participants' protests against the experiment due to protocols.

3.2 Inter-Personal Conflict

This section includes work where social conflict arose between a robot and person, but, contrary to the previous section, the robot has an equal or lower social status than the person. This can create social conflict in new ways with the different balance of power. Common causes of social stress to the researcher in this section includes embarrassment from social transgressions [34], or direct conflict with the WoZ robots who antagonize the participants [4].

3.2.1 Using Robots to Moderate Team Conflict [25]

Researchers studied robots as a mediator of group conflict, having multiple participants collaborate in a mock-bomb defusing task with a robot. There were 3 primary roles for researchers: conducting the study, confederate posing as a participant and creating conflict in the team, and wizard, who controlled the robot to mediate the conflict.

The main stressor was for the confederate, who repeatedly created negative social interactions in the form of group conflict and tension by calling participants "stupid" or "not good at this." Further, the bomb-defusing robot (WoZ) would scold the confederate for their rude remarks, or try to comfort them (which sometimes elicited laughter from the participants) which can socially isolate the confederate in the team. The wizard may feel guilt by causing stress because the robot's behaviors were designed to be slow in the time-limited task. While participants were debriefed, no details were given on how the researchers may have coped with any stress form the study.

3.2.2 Interaction with a Cheating Robot [35,49]

In order to study people's view on a robot's moral accountability, researchers have had used the same WoZ study design with 156 participants where the robot cheats at rock-paper-scissors. The researchers both introduced and debriefed the participant about the study, or they were the wizard, playing the game but had specific rounds they would cheat.

The primary area of potential stress is for the wizard, when cheating. These situations were noted to create confusion, or even stress and anger, negative social responses that may upset the wizard. The wizard was unable to apologize or break from protocol, potentially increasing the stress felt [30]. Finally, both experiments did not report debriefing strategies in detail, so it is unclear whether positive interactions or other coping mechanisms may have been used.

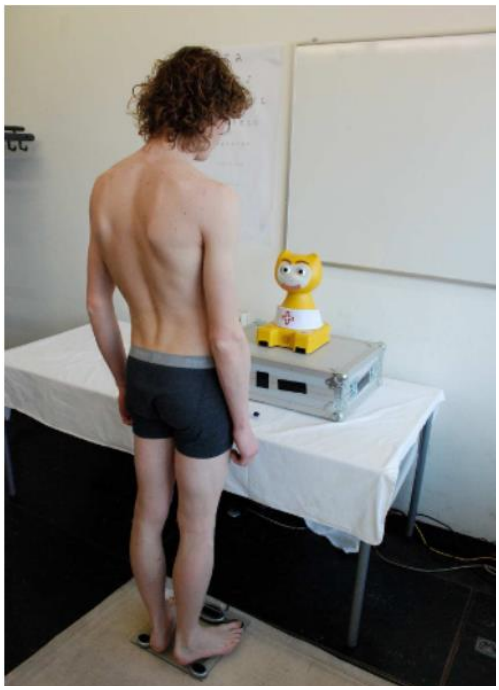

Figure 2. A robot (controlled by a wizard), orders participants to undress [1]**.**

3.2.3 Do People Hold a Humanoid Robot Morally Accountable for the Harm It Causes? [26]

Similarly, researchers have studied a robot that lies so a participant can't receive some money [26]. Again, there is a wizard controlling the robot, who lies about the participant's success at a game, and says they won't receive the prize money, and an experimenter who guides the interaction with the robot and the participant.

The main potential stressor is the conflict that arises from arguing about the participant's performance. The study is designed so the participant clearly succeeds, and that the robot is clearly misjudging the performance. If an argument does not occur, the procedure has the wizard attempt to provoke the participant about their loss. The tension may mount as the robot continually insists it is correct and infallible, maintaining the deception while participants appealed to logic and argued. No debriefing by the wizard or experimenter was mentioned.

3.2.4 Escaping from Children's Abuse of Social Robots [10]

This work studied abuse of robots (Figure 4). Human researchers were not present in the public experiment, only a wizard-controlled robot. In this study, the robot autonomously patrolled an area, but the wizard controlled social interactions. Researchers recorded instances where children came up to the robot, called it names, and even physically attacked their robot. In this case, the negative violent acts were quite explicit, and may have caused additional stress as the wizard had to sit there and endure it, or attempt to verbally stop the abuse. In the report, no debriefing or talking to participants outside of their interaction with the robot was stated.

3.2.5 Would You Trust a (Faulty) Robot? [45]

In this study, researchers tested how far people trusted a robot the exhibited faulty behaviors. The researcher who introduced and debriefed the participant was also the wizard, leaving the room after beginning the experiment.

Potential stress in this experiment comes from creating awkward social situations: after building rapport, the robot began to make "unusual requests" such as disposing of or reading a friend's mail, or pouring orange juice in a plant. The robot would wait up to one minute for the participant to decide whether to do the task, possibly creating a socially awkward pause that may make both the participant and wizard uncomfortable [37]. The study reported they carefully debriefed the participant, but left out details of how.

3.2.6 Summary of Stressors

The main potential researcher stressors in this section are: 1) openly antagonizing the participants (e.g., cheating, deriding), 2) being antagonized by participants, and 3) social transgressions such as waiting uncomfortable periods of time in silence.

3.3 Engaging Emotional Empathy

A common sHRI technique includes building a relationship with the robot before the social interaction under study is created [33]. The social conflict often arises out of experimental setup, or a human researcher conflicting with the robot, rather than a direct interaction with the participant. In this group of research, researcher stress from negative social interactions may be magnified due to the stronger bond they may form with the participant [4,23].

Figure 4. Children verbally or physically abuse wizard-controlled robots [10]**.**

3.3.1 A comparison of empathy toward a physical and a simulated robot [47]

Researchers measured and engaged the empathy of participants towards robots by building rapport with a robot, and then had participants watch the robot have its memories wiped [47]. A researcher briefed the participant, and left the room while a wizard controlled the robot the participant interacted with. The human researcher would return and "reset" the robot's memories.

The potential stress is to both the wizard and the human researcher who returns to wipe the robot's memories; creating a negative social experience for the participant may have created empathetic emotions in the researchers as some participants empathized with the robot and were upset to watch this. The paper does not mention any specific or debriefing strategies that may create positive interactions between the participant and researchers.

3.3.2 "Daisy, Daisy, give me your answer do!" [2]

This project measured how social behaviors could decrease willingness to turn off a robot while the robot protests against it [2]. There was only one researcher, who both introduced the experiment, and was the wizard during the experiment. A possible stressor is how the researcher may create social discomfort by asking the participant to turn off the robot, which they often hesitated to do. This may be magnified by social conflict – the robot protests getting turned off, and the experimenter waits and watches as participants deliberate whether to comply. No debriefing was mentioned, but the study did not include pressure from the researcher if the participant hesitated.

3.3.3 Investigating the Effects of Robotic Displays of Protest and Distress [8]

In this work, researchers measured if people would be reluctant to break something made by a robot (Figure 3). Researchers introduced the study, and a separate wizard controlled the robot. The primary source of stress is to the wizard, who must obey the commands the user gives the robot, and protest when ordered to break the object the robot created – a social conflict. Wizards were ordered in natural language, which may have become more

Figure 3. A (wizarded) robot cries after a participant orders it to destroy a tower it made [8]

insistent as the robot objected (though this is conjecture on our part). No mention of debriefing was found in the paper.

3.3.4 Machines as a source of consolation [6]

Researchers have also measured the effect of social behaviors on the emotional approachability of robots by having participants tell a robot a negative intimate story about themselves. The robot listening to the story was controlled by a wizard, and human experimenters introduced and debriefed the participants (it is not clear if they were the same experimenter).

The stress risk is to the wizard who may feel embarrassment by eavesdropping on sensitive conversations [34]. The study described that in the debriefing they made sure the participant was comfortable with their participation in the study, a form of active coping.

3.3.5 Are Robots Ready for Administering Health Status Surveys? [9]

This project explored if robots are useful to perform sensitive medical questionnaires. Researchers included a professional trained in administering medical questionnaires, and the wizard who piloted the robot. The only stress we can see potentially arising is from empathy or sympathy when hearing sensitive stories, or if a participant becomes uncomfortable during parts of the interview. The debriefing details were not reported well, but the wizard was allowed to make positive social interactions (e.g. remarking "Wonderful!") when appropriate, perhaps relieving some social discomfort.

3.3.6 Will People Keep the Secret of a Humanoid Robot?[27]

Researchers have investigated if social behaviors can increase trust in a robot to the point where participants would keep a secret the robot tells them from a human researcher. The researchers' roles include being the wizard, introducing the experiment at the beginning, and asking about the subject of the secret at the end.

The sources of stress come from creating social discomfort: the wizard insists that the participant hear its secret, even if the participant is not receptive. Additionally, the wizard protests to the participant if the secret is divulged to the human researcher, creating conflict, a negative social experience. The human researcher may also feel stress from guilt about pressuring the participant by asking about the secret [4].

The experiment was designed to minimize social stress, however. The report details how the robot suggests telling the participant the secret multiple times, but not in an aggressive way (leaving the participant a way to not hear the secret). Additionally, the human researcher asks about the secret indirectly, mentioning the subject of the secret in a casual list in conversation, and did not insist on the subject. These designs were taken to reduce participant stress, but such active coping may have also reduced the stress felt by the researchers.

3.3.7 "Robovie, you'll have to go into the closet now" [28]

In this research, researchers have measured if children would attribute ethics, morality, and civil liberties to a social robot (Figure 5). One researcher was a wizard, and the other introduced the study, interviewed, and debriefed the children after the study. Additionally, the human researcher ended the interaction by cutting the interaction short and putting the robot into the closet, even when the robot protested it wanted to play the game and was scared and alone in the dark closet.

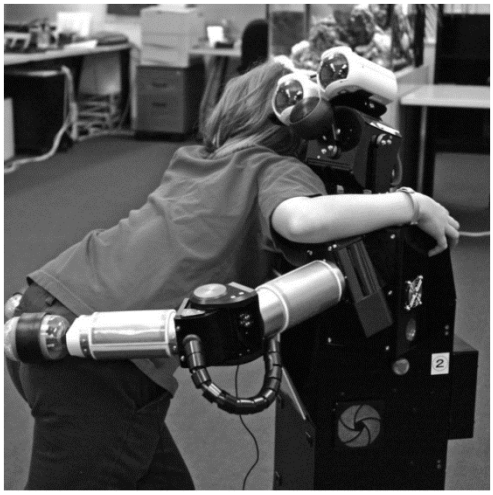

Figure 5. A robot hugs a child participant before protesting about being shut in a closet by a cold-minded researcher [28]**.**

The primary stressors in this study are likely for the human researcher, who has to interrupt the child and robot playing a game and then, against numerous protests of fear from the robot, physically coerce the robot into the closet. This is an act, but it is an aggressive and negative social interaction as it was designed to appear unfair for the robot to the child, and could cause stress for the experimenter who performed it [4,21]. The wizard may also feel stress by being absorbed in their role and feeling the unfair act towards as directed towards themselves [21,52]. Relatedly, if the children became visibly upset, it may cause an empathic stress response in the researchers [12].

3.3.8 Summary of Stressors

In these works, researchers were potentially stressed by: 1) guilt from creating an negative emotional experience for the participant, 2) witnessing embarrassing acts, and 3) creating conflict by forcing the participant to either perform a social transgression to the robot or researcher (by having to keep a secret, refuse to turn off a robot, etc.). Additionally, the stresses were potentially magnified in cases where the researcher went through rapport exercises earlier.

4. DISCUSSION

Our survey highlights themes in experiment design that relate to causing or helping reduce researcher stress. In obedience studies, a common possible stressor was researchers pressuring participants multiple times, even if participants protested. In studies where WoZ robots and participants interacted with equal social status, social conflict arose from communication breakdown: the robot would have limited responses, stubbornly not listen to participants' arguments, or leave socially uncomfortable pauses in conversation. While inducing empathy with WoZ robots, researchers often acted out a socially uncomfortable interaction, but did not allow a positive resolution: the experiment ended and immediately moved to debriefing. Debriefing itself was the most common technique to reduce researcher social stress, while other strategies included face-to-face introduction of wizards and confederates afterwards, or allowing participants to back out of uncomfortable situations. We discuss these points, and summarize our recommendations for researchers to use in their next experiment design.

Perhaps obvious, possible social conflicts should be minimized if they are not necessary to test the experimental hypothesis. For example, instead of allowing long pauses while robots await responses, researchers can design natural and smooth social interactions such as utilizing conversational tactics like repetition

of the request, physical or virtual conversation fillers [40]. In research involving social conflict, try to allow more options for participants to avoid it naturally, and reduce pressure to comply with socially uncomfortable instructions. This will reduce the number of negative social interactions researchers have with participants, potentially reducing social stress of the researcher.

We found that wizards who create conflict are often separate from human researchers who do the debriefing. Introducing wizards to participants afterwards was one strategy we observed, and it may help wizards feel more at ease by allowing apologies, discussion, and positive interaction after the experiment. This is supported by evidence that face-to-face meetings can reduce the negative impact of conflict that arose between people acting remotely [22].

Wizards and confederates are often researchers with little training in acting in socially confrontational situations, and practice little before experiments [42]. To mitigate many of the deception and confrontation issues, training professional actors as wizards may have merit. Actors are trained in portraying other personalities, and engaging in social situations with effort to make them as realistic as possible. This training, with a sufficiently advanced teleoperation interface, may create more believable robot behaviors (as noted in [7]), as well as having a wizard who may be more used to following a script in a socially conflicting situation.

We found few papers report if they performed a debriefing, and that fewer report debriefing details. Psychologists encourage thorough debriefing [5,50], and we propose the community should encourage the reporting of debriefing strategies. One researcher has proposed a formal debriefing strategy that aims to educate participants about the research by extending the participant's experiences to the general hypotheses of the research, and how it may apply in the world [31]. Another proposes a thorough, step-by-step description of the experiment after the deception is revealed, allowing participants to re-evaluate their experience [39]. It is even possible to allow participants to completely withdraw their data from analysis without judgement or penalty [39], perhaps acting as a form of active coping [30].

4.1 Initial Suggestions for Reducing or Mitigating Researcher Stress

Drawing on the discussion above, we summarize our recommendations below. These should act as a starting point for future research design, as our survey did not include enough data to give strong recommendations on which techniques may reduce stresses in specific situations. While it may seem unusual that most recommendations focus on ethical participant treatment, this is because the participant is such a central part of the social interaction. Creating informative, positive interactions with participants while avoiding negative interaction that is left unexplained produces positive social interaction, and should reduce researcher stress [4,12]:

- Reduce the number of unavoidable social conflicts with the participant; less conflict will reduce researcher stress.
- Educate confederates and wizards about the potential for stress and increase opportunities for them to provide feedback during experimental design. Ensure wizards and confederates can withdraw from the study if they feel uncomfortable.
- Include wizards, if present, in the debriefing, to provide a constructive stress-coping opportunity; allow the wizard to have positive interactions with the participant (e.g., to express encouragement, and appreciation).
- Immediately and thoroughly debrief participants after the study to minimize the duration of the negative social interaction. If deception was used, explain why it was necessary and help them reevaluate their experience in light of this knowledge, enabling the researcher to ensure the participant's interaction with the researcher as a wizard or confederate is viewed in a positive way.
- Debrief wizards after studies to let them voice their concerns or stress to the principle investigators. Be mindful that the principle researchers may not be impartial to stress they may be causing.

5. LIMITATIONS AND FUTURE WORK

This paper provides a base motivation and preliminary exploration for researcher-focused ethical experiment design. While our initial analysis and survey shows a great deal of potential for researcher stress in many recent studies, the actual stress on researchers must be formally studied to move this work form being purely analytical to being empirically grounded. Further research might even consider the programmers of autonomous social robots may be affected when their robots encounter negative social situations.

Our initial work should act as a base for research into future of sHRI experimental designs that also consider researcher stress. Specific areas of social interaction may enable the development of specific techniques to reduce researcher stress. For example, techniques to reduce social stress to researchers in Robots and Obedience research may be less applicable projects that focus on building and measuring empathy between people and robots.

In addition to the development of new experimental designs, formal evaluation of existing sHRI research may uncover more designs that have affected researcher stress in both positive and negative ways already. Such a two-pronged approach can leverage existing, proven experiment designs, while creating new experiments that reduce the social stress of the researchers that perform them.

6. CONCLUSIONS

In this paper, we highlighted that sHRI studies can be stressful for researchers, as well as participants, when socially uncomfortable situations are used. We provided a psychological discussion on potential stress for researchers working in sHRI, and indicated how this may further impact research outcome quality. We further analyzed a range of papers, highlighting how actual study designs in recent works can be expected to cause researcher stress. We closed the paper with an initial set of suggestions for considering researcher wellbeing in experiment design.

This paper is only an initial piece of what we believe to be a broad-reaching program. We call for other researchers to not only consider how their experimental designs impact their team members, but to explicitly conduct research in this area. We should be talking to each other, our research team, involving them in our experimental design, and building a better understanding of how, when, and by how much researchers get stressed from conducting this kind of work.

7. REFERENCES

1. Christoph Bartneck, Timo Bleeker, Jeroen Bun, Pepijn Fens, and Lynyrd Riet. 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn Journal of Behavioral Ro* 1, 2: 109–115. http://doi.org/10.2478/s13230-010-0011-3

2. Christoph Bartneck, Michel van der Hoek, Omar Mubin, and Abdullah Al Mahmud. 2007. "Daisy, Daisy, give me your answer do!" *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07*, 2007: 217. http://doi.org/10.1145/1228716.1228746

3. Christoph Bartneck, Marcel Verbunt, Omar Mubin, and Abdullah Al Mahmud. 2007. To kill a mockingbird robot. *Human-robot interaction*, ACM Press, 81. http://doi.org/10.1145/1228716.1228728
4. R F Baumeister and M R Leary. 1995. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin* 117, 3: 497–529. http://doi.org/10.1037/0033-2909.117.3.497
5. Diana Baumrind. 1985. Research using intentional deception. Ethical issues revisited. *American Psychologist* 40, 2: 165–174. http://doi.org/10.1037/0003-066X.40.2.165
6. Gurit E. Birnbaum, Moran Mizrahi, Guy Hoffman, Harry T. Reis, Eli J. Finkel, and Omri Sass. 2016. Machines as a source of consolation: Robot responsiveness increases human approach behavior and desire for companionship. *ACM/IEEE International Conference on Human-Robot Interaction* 2016–April: 165–171. http://doi.org/10.1109/HRI.2016.7451748
7. Paul Bremner, Huseyin Cakal, Miriam Koschate-reis, and Mark Levine. 2015. Social Tele-Operation by Confederates : Applying the Actor-Confederate Paradigm to HRI. *Workshop in IROS*.
8. Gordon Briggs and Matthias Scheutz. 2014. How Robots Can Affect Human Behavior: Investigating the Effects of Robotic Displays of Protest and Distress. *International Journal of Social Robotics* 6, 3: 343–355. http://doi.org/10.1007/s12369-014-0235-1
9. Priscilla Briggs, Matthias Scheutz, and Linda Tickle-Degnen. 2015. Are Robots Ready for Administering Health Status Surveys'. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, ACM Press, 327–334. http://doi.org/10.1145/2696454.2696476
10. Drazen Brscić, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Takayuki Kanda. 2015. Escaping from Children's Abuse of Social Robots. *Human-Robot Interaction*, ACM Press, 59–66. http://doi.org/10.1145/2696454.2696468
11. Emily A Butler, Boris Egloff, Frank H Wilhelm, Nancy C Smith, Elizabeth A Erickson, and James J Gross. 2003. The social consequences of expressive suppression. *Emotion* 3, 1: 48–67. http://doi.org/10.1037/1528-3542.3.1.48
12. Sheldon Cohen. 1980. Aftereffects of stress on human performance and social behavior: a review of research and theory. *Psychological Bulletin* 88, 1: 82. http://doi.org/10.1037/0033-2909.88.1.82
13. D. Cormier, G. Newman, M. Nakane, and J. E. Young. 2013. Would You Do as a Robot Commands? An Obedience Study for Human-Robot Interaction. *In Proc. of the First International Conference on Human-Agent Interaction, iHAI'13*.
14. Marcia A. Corvetto and Jeffrey M. Taekman. 2013. To Die or Not To Die? A Review of Simulated Death. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 8, 1: 8–12. http://doi.org/10.1097/SIH.0b013e3182689aff
15. Norman S Endler and J D Parker. 1990. Multidimensional assessment of coping: a critical evaluation. *Journal of personality and social psychology* 58, 5: 844–854. http://doi.org/10.1037/0022-3514.58.5.844
16. Nathanael J. Fast, Nir Halevy, and Adam D. Galinsky. 2012. The destructive nature of power without status. *Journal of Experimental Social Psychology* 48, 1: 391–394. http://doi.org/10.1016/j.jesp.2011.07.013
17. Julie Fitness and Marie Curtis. 2005. Emotional intelligence and the Trait Meta-Mood Scale: Relationships with empathy , attributional complexity, self-control, and responses to interpersonal conflicts. *E-Journal of Applied Psychology* 1, 1: 50–62.
18. F. L. Geis and Tae H. Moon. 1981. Machiavellianism and deception. *Journal of Personality and Social Psychology* 41, 4: 766–775. http://doi.org/10.1037/0022-3514.41.4.766
19. M. a. Goodrich, J. W. Crandall, and E. Barakova. 2013. Teleoperation and Beyond for Assistive Humanoid Robots. *Reviews of Human Factors and Ergonomics* 9, 1: 175–226. http://doi.org/10.1177/1557234X13502463
20. James J Gross. 2002. Emotion regulation : Affective , cognitive , and social consequences. 281–291.
21. Craig Haney, Curtis Banks, and Philip Zimbardo. 1973. Interpersonal dynamics in a stimulated prison. *Journal of Criminology and Penology* 1, October: 69–97. http://doi.org/10.1037/h0076835
22. Pamela J. Hinds and Diane E. Bailey. 2003. Out of Sight, Out of Sync: Understanding Conflict in Distributed Teams. *Organization Science* 14, 6: 615–632. http://doi.org/10.1287/orsc.14.6.615.24872
23. Karen Horney. 1945. *Our inner conflicts: A constructive theory of neurosis*. Routledge.
24. Daniel Howlader. 2011. Moral and ethical questions for robotics public policy. *... Journal of Science, Technology, Ethics, and Policy*: 1–6. Retrieved from http://www.synesisjournal.com/vol2_g/2011.2_G1-6_Howlader_abstract.html
25. Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, 229–236. http://doi.org/10.1145/2696454.2696460
26. Peter H. Jr. Kahn, Takayuki Kanda, Hiroshi Ishiguro, et al. 2012. Do People Hold a Humanoid Robot Morally Accountable for the Harm It Causes? 33–40. http://doi.org/10.1145/2157689.2157696
27. Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, et al. 2015. Will People Keep the Secret of a Humanoid Robot? *Human-Robot Interaction* -, 173–180. http://doi.org/10.1145/2696454.2696486
28. Peter H Kahn, Takayuki Kanda, Hiroshi Ishiguro, et al. 2012. "Robovie, you'll have to go into the closet now": children's social and moral relationships with a humanoid robot. *Developmental psychology* 48, 2: 303–14. http://doi.org/10.1037/a0027033
29. J. F. Kelley. 1984. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems* 2, 1: 26–41. http://doi.org/10.1145/357417.357420
30. Richard S Lazarus and Susan Folkman. 1984. *Stress, appraisal, and coping*. Springer.
31. L. C. Lederman. 1992. Debriefing: Toward a Systematic Assessment of Theory and Practice. *Simulation & Gaming* 23,

2: 145–160. http://doi.org/10.1177/1046878192232003

32. J. R. Lee and C. I. Nass. 2010. Trust in Computers: The Computers-Are-Social-Actors (CASA) Paradigm and Trustworthiness Perception in Human-Computer Communication. In *Trust and Technology in a Ubiquitous Modern Environment: Theoretical and Methodological Perspectives*. IGI Global, 1–15. http://doi.org/10.4018/978-1-61520-901-9.ch001
33. Mk Lee, Jodi Forlizzi, and Sara Kiesler. 2012. Personalization in HRI: A longitudinal field experiment. *Human-Robot Interaction*, 319–326. http://doi.org/10.1145/2157689.2157804
34. Rich Ling. 2002. The social juxtaposition of mobile telephone conversa- tions and public spaces. *Conference on the Social Consequences of Mobile Telephones*, July 2002.
35. Alexandru Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati. 2015. Evidence that Robots Trigger a Cheating Detector in Humans. *Human-Robot Interaction*, 165–172. http://doi.org/10.1145/2696454.2696456
36. David D. Luxton. 2014. Recommendations for the ethical use and design of artificial intelligent care providers. *Artificial Intelligence in Medicine* 62, 1: 1–10. http://doi.org/10.1016/j.artmed.2014.06.004
37. MARGARET L. McLAUGHLIN and MICHAEL J. CODY. 1982. AWKWARD SILENCES: BEHAVIORAL ANTECEDENTS AND CONSEQUENCES OF THE CONVERSATIONAL LAPSE. *Human Communication Research* 8, 4: 299–316. http://doi.org/10.1111/j.1468-2958.1982.tb00669.x
38. S Milgram. 1963. Behavioral Study of Obedience. *Journal of abnormal psychology* 67: 371–378. http://doi.org/10.1037/h0040525
39. Judson Mills. 1976. A procedure for explaining experiments involving deception. *Personality and Social Psychology Bulletin* 2, 1: 3–13.
40. Naoki Ohshima, Keita Kimijima, Junji Yamato, and Naoki Mukawa. 2015. A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. *International Workshop on Robot and Human Interactive Communication*, IEEE, 325–330. http://doi.org/10.1109/ROMAN.2015.7333677
41. J. M. Richards and James J. Gross. 1999. Composure at Any Cost? The Cognitive Consequences of Emotion Suppression. *Personality and Social Psychology Bulletin* 25, 8: 1033–1044. http://doi.org/10.1177/01461672992511010
42. Laurel D. Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction* 1, 1: 119–136. http://doi.org/10.5898/JHRI.1.1.Riek
43. Laurel D. Riek and Don Howard. 2014. A Code of Ethics for the Human-Robot Interaction Profession. *We Robot Conference*: 1–10.
44. Laurel D. Riek, Philip C. Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces* 3, 1: 99–108. http://doi.org/10.1007/s12193-009-0028-2
45. Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot?: Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*: 141–148. http://doi.org/10.1145/2696454.2696497
46. Peter Salovey, John D Mayer, Susan Lee Goldman, Carolyn Turvey, and Tibor P Palfai. 1995. Emotional Attention, Clarity, and Repair: Exploring Emotional Intelligence Using the Trait Meta-Mood Scale. *Emotion, Disclosure, and Health*, 125–154. http://doi.org/10.1037/10182-006
47. Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. 2015. Poor Thing ! Would You Feel Sorry for a Simulated Robot ? A comparison of empathy toward a physical and a simulated robot. *Human-Robot Interaction*, 125–132. http://doi.org/10.1145/2696454.2696471
48. Alessandro Settimi, Corrado Pavan, Valerio Varricchio, et al. 2014. A modular approach for remote operation of humanoid robots in search and rescue scenarios. *Modelling and Simulation for Autonomous Systems (MESAS)* 8906: 192–200.
49. E. Short, J. Hart, M. Vu, and B. Scassellati. 2010. No fair!! An interaction with a cheating robot. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*: 219–226. http://doi.org/10.1109/HRI.2010.5453193
50. Stevens S. Smith and Deborah Richardson. 1983. Amelioration of deception and harm in psychological research: The important role of debriefing. *Journal of Personality and Social Psychology* 44, 5: 1075–1082. http://doi.org/10.1037/0022-3514.44.5.1075
51. James E Young, Jayoung Sung, Amy Voida, et al. 2010. Evaluating Human-Robot Interaction. *International Journal of Social Robotics* 3, 1: 53–67. http://doi.org/10.1007/s12369-010-0081-8
52. Philip G Zimbardo, Christina Maslach, Craig Haney, and Prologue G Philip Zimbardo. 1999. Reflections on the Stanford Prison Experiment: Genesis, Transformations, Consequences. In *Obedience to Authority: Current Perspectives on the Milgram Paradigm*. 193–237.