MASTERARBEIT | MASTER'S THESIS

Titel | Title

The Interplay of Decision Agency, Outcome Evaluation and Outcome Ownership When Interacting with Robots
Hands Off, Robots!

verfasst von | submitted by

Anja Huber

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 013
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Joint-Degree Middle European interdisc. master prog. in Cognitive Science
Betreut von | Supervisor:	Dr. Simone Schütz-Bosbach

I am grateful to:
Simone Schütz-Bosbach
Maren Giersiepen
Tomas Weinert
Jakob Kaiser

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

In contemporary society, the emergence of social and intelligent robots has become increasingly prevalent (Broadbent, 2017). However, research suggests that the interaction with robots can lead to cognitive conflict in humans (Abubshait et al., 2023). One aspect that may influence this conflict is the degree to which a task is outsourced to the robot. This study aims to expand upon existing knowledge (Abubshait et al., 2023) regarding the influence of outcome value (positive vs. negative outcome) and outcome ownership (giving the outcome to the robot vs. keeping the outcome to oneself) on cognitive conflict when interacting with robots. To this regard, different levels of decision-making-agency (self-determined choice vs. externally determined choice) were incorporated into the investigation. To address this question, an online experimental framework was employed, in which participants engaged in interactive tasks with a social intelligent robot named Cozmo, materialized as an avatar and introduced in the beginning of the experiment. The participant or the robot chose between two presented stimuli, resulting in either a win or loss of a small amount of money, which was then assigned to either the participant or the robot. We assessed cognitive conflict by examining the participants' response times when confirming outcome attributions for each trial, with longer response times indicating higher levels of cognitive conflict.

Our findings align with previous research by Abubshait et al. (2023), who observed that participants exhibited longer response times when assigning outcomes to the robot. Additionally, we found that decision agency significantly influenced response times and interacted with the other variables under investigation. Specifically, participants responded faster, suggesting lower cognitive conflict, in scenarios where they take the decision themselves, as well as when they self-own a positive outcome. We think that our findings will contribute to a better understanding of how to best design and implement socially intelligent robots for fruitful human-robot interaction.

KURZREFERAT

In unserer heutigen Gesellschaft sind soziale und intelligente Roboter zunehmend verbreitet (Broadbent, 2017), jedoch deuten Forschungsergebnisse darauf hin, dass die Interaktion mit Robotern bei Menschen zu kognitiven Konflikten führen kann (Abubshait et al., 2023). Eine potenzielle Variable, die auf diesen Konflikt einwirken kann, ist das Ausmaß, in dem eine Entscheidungs-Aktions-Situation an den Roboter ausgelagert werden. Die hier beschriebene Studie zielt darauf ab, das bereits bestehende Wissen von Abubshait et al. (2023) über den Einfluss des Ergebniswertes (positives vs. negatives Ergebnis) und der Ergebniszuschreibung (das Ergebnis dem Roboter geben oder das Ergebnis für sich behalten) auf den kognitiven Konflikt bei der Interaktion mit Robotern zu erweitern, indem verschiedene Ebenen der Entscheidungsbefugnis (selbstbestimmte Wahl vs. fremdbestimmte Wahl) in die Untersuchung mit einbezogen werden. Um diese Frage zu klären, wurde ein Online-Experiment erhoben, bei dem die Teilnehmenden interaktive Aufgaben mit einem sozialen Roboter namens Cozmo bearbeiteten, der als Avatar dargestellt und zu Beginn des Experiments vorgestellt wurde. Die Teilnehmenden oder der Roboter wählten zwischen zwei dargebotenen Stimuli, die entweder zu einem Gewinn oder Verlust eines kleinen Geldbetrages führten, der dann entweder dem Teilnehmenden oder dem Roboter zugewiesen wurde. Als Maß für den kognitiven Konflikt untersuchten wir die Reaktionszeiten in der Bestätigung der Ergebniszuschreibung jedes Trials, wobei wir verlängerte Reaktionszeiten als Hinweis auf ein erhöhtes Maß an kognitivem Konflikt interpretierten.

Unsere Ergebnisse bestätigen den bereits in früheren Untersuchungen von Abubshait et al. (2023) gefundenen Haupteffekt, welcher signifikant erhöhte Reaktionszeiten der Teilnehmenden belegte, wenn diese die Gewinne oder Verluste an den Roboter abgeben mussten. Des Weiteren konnten unsere Analysen zeigen, dass die Entscheidungsbefugnis signifikanten Einfluss auf die Antwortzeiten hatten und dass ein Interaktionseffekt zwischen allen drei untersuchten Variablen bestand. Die Teilnehmenden erreichten kürzere Reaktionszeiten in Trials, in denen sie die Entscheidung selbst treffen konnten und wenn diese ein positives Ergebnis selbst behalten durften. Dies deutet auf einen geringen kognitiven Konflikt in solchen Szenarien hin.

Wir denken, dass unsere Ergebnisse zu einem besseren Verständnis darüber beitragen, wie sozial intelligente Roboter gestaltet werden müssen, um eine effektive Interaktion zwischen Menschen und Robotern zu gewährleisten.

I INTRODUCTION

In our rapidly evolving world, technological progress is accelerating and fundamentally changing our daily lives. From the prevalence of smartphones to the rise of smartwatches, the way we interact with technology is constantly changing. Technological innovations such as self-service checkouts, autonomous vehicles, parcel lockers replacing traditional post offices and massage robots are becoming an integral part of our daily routine. Particularly striking is the increasing presence of robots in various sectors including healthcare, where robot assistance is becoming more common. As we look to the future, it is obvious that robots will be used in an ever-increasing number of applications. From housework to complex medical procedures, their tasks will become more diverse and promise a world in which automation and artificial intelligence are deeply embedded in our everyday experiences.

However, these developments are not without challenges and concerns. The widespread adoption of robots brings significant fears and ethical considerations with it. The potential use of robots in warfare, the risk of dehumanization and the possibility of diminished human interaction are pressing issues that cannot be ignored. These developments raise important questions about the implications of a world increasingly populated by autonomous machines. Current research therefore focuses not only on how we interact with robots, but also on how they should be designed and implemented to ensure that these interactions are seamless and beneficial (Abubshait et al., 2023). Understanding the dynamics of human-robot interaction is critical for developing technologies that integrate smoothly into our lives, promote acceptance and minimize discomfort. The goal is to develop robots that are not only functional, but also improve human quality of life.

This thesis delves into questions related to the design and implementation of robots and the evolving relationship between humans and machines. It will deepen our understanding of the dynamics underlying human-robot interaction, shedding light on the nuanced interplay of decision agency, outcome evaluation and outcome ownership in decision action situations when interacting with robots.

II THEORETICAL BACKGROUND

2.1 Prosocial Behavior Towards Robots

Over the last 30 years, we have seen how intelligent robots and other modern technologies are progressively shaping our everyday lives. Even if these robots have not yet achieved the functionality and intelligence that are portrayed in many science fiction films (Broadbent, 2017), it is likely that people are transferring the concerns they have about digital devices (Chin et al., 2012) to modern social robots (Broadbent, 2017). This trend is worrying because the use of robots promises to be of great help in solving social problems such as caring for the ever-increasing proportion of the world's aging population (Broadbent, 2017).

Current research projects, which are carried out with both, real robots as well as simulations of them (the latter using only "written descriptions, photographs, or videos of robots", see Broadbent, 2017, p. 637) often apply traditional psychology research paradigms to find differences between behavior towards robots in comparison to behavior towards other human beings (Broadbent, 2017). Although people tend to attribute human characteristics to robots and, as a result, treat them as humans in many situations, it appears that there are very nuanced situational factors that can influence how humans behave towards robots (Broadbent, 2017). For instance, Nass et al. (1997) could show that people apply the same strong gender bias when using computers as they apply towards other humans. Among other things, their study confirmed the well-known stereotype that when being evaluated by a male voice, humans consider the given feedback as more valid than the one by a female voice. Nass et al. (1997) state that „for designers and engineers to assume that any voice is neutral, is a mistake; a male voice brings with it a large set of expectations and responses based on stereotypes about males, whereas a female voice brings with it a large set of expectations and responses based on stereotypes about females" (p. 11). Another impressive example in which human interaction with robots differs from that with other humans comes from Malle et al. (2015). With their experiment, they were able to show that humans tend to apply different moral standards to robots as compared to other humans. In the narrative scenarios they presented, participants were more likely to expect utilitarian decisions from robots, such as sacrificing a human to save others, than they did from humans (Malle et al., 2015).

> [S]ocial robots are starting to become more common in our society and can benefit us by providing companionship, increasing communication, and reducing costs, especially in healthcare settings. …. However, robots can also make us feel uncomfortable, especially when their appearance is inconsistently humanlike and thus threatening or when their behavior violates expectations. Interacting with robots can make us question who we are, and we can project our desires onto them. Theories suggest we are wired through a combination of nature and nurture to perceive robots through human filters. We mindlessly interact with robots and other technologies as if they were human and we perceive humanlike characteristics in them, including thoughts and emotions. However, we do not see them as having as much mind as humans do, nor do we ascribe to them the same moral rights and responsibilities as humans. (Broadbent, 2017, p. 645)

With this statement, Broadbent emphasizes that there are many nuanced differences in how we interact with or treat non-human agents and that, due to the high potential artificial agents offer for the future, society should carefully consider and research how to best design and implement robots into our lives. This research project follows this thought by aiming to expand our knowledge about pro-social behavior towards robots.

2.2 Cognitive Dissonance

When the interaction with robots is discussed in society, concerns regarding "deception, privacy, job loss, safety, and the loss of human relationships" (Broadbent, 2017, p. 627) are frequently mentioned. When assuming a real-life interaction between a human and a robot, those concerns might manifest themselves in humans as cognitive dissonance, a concept first introduced by Leon Festinger in 1957 (Festinger, 2001). Festinger (2001) defines cognitive dissonance as an uncomfortable state of the mind emerging through contradicting cognitions. His theoretical framework explains how cognitive conflict in humans can arise and suggests that cognitions (such as beliefs or attitudes) can be irrelevant or relevant to one another (Festinger, 2001). Relevant cognitions are those that are thematically related, such as the thought that one wants to live healthily and the perception of fresh vegetables during the preparation of a meal. Irrelevant cognitions are those that are not related, such as the thought of wanting to live healthily and the perception of the color of the wall opposite oneself. If cognitions are considered relevant by an individual, then there is the option for the cognition to be ‘consonant’ or ‘dissonant’. Consonant cognitions are those who follow from each other, fit together, and do not contradict one another (Festinger, 2001). For instance, the intention to save money and noticing that the balance of your bank account is increasing every month can be considered a consonant cognition. Dissonant cognitions are those that are relevant but contradict each other (Festinger, 2001). For example, the intention to save money and noticing that the balance of your bank account is decreasing. Cognitions which are relevant and dissonant are leading to the state of cognitive dissonance also called cognitive conflict (Harmon-Jones & Mills, 2019). In the context of human-robot interactions, this conflict may arise from the necessity to collaborate with robots to complete a task, despite a reluctance or discomfort in doing so.

When a person experiences cognitive dissonance, he or she would usually try to reduce the dissonance by either changing his or her behavior or cognition. Festinger's theory (2001) has far-reaching positive and negative consequences on how one can interpret topics of everyday life and society. One example for its application on current societal issues is individual media consumption. It has been confirmed that people tend to consume media that corresponds to their existing opinions (Jean Tsang, 2019), according to Festinger’s theory (2001) this is the case because less dissonance arises from content in line with own cognitions. This goes hand in hand with people putting more trust in this type of media than in reports that contradict their own views (Metzger et al., 2020). In recent years, the polarization of society has been a frequent topic of discussion. Although it is not clear whether society is more polarized today than in the past (Mau et al., 2024), media consumption undoubtedly influences people's opinions. Over time, a selective exposure to media might contribute to people becoming entrenched and more extreme in their views, which further increases the division of society. But also positive examples of how Festinger’s dissonance theory (2001) explains individual behavior in important societal issues can be found. Research has shown how the application of the concept can be used to convince people to donate for homeless (Fointiat, 2004), support water conservation measures (Fointiat, 2004), or motivate young adults to use contraceptives (Stone et al., 1994).

Current research suggest that discomfort is also perceived when collaborating or interacting with robots (Levin et al., 2013; Abubshait et al., 2023). Levin et al. (2013) let their participants perform a medical triage situation either with a human partner or a robot partner, where both potential partners provided the same guidance regarding steps to perform in the triage and further recorded the performed measures by the participant and results of the dummy patient.

They found that in the condition in which humans were paired with robots, participants experienced higher levels of cognitive dissonance, compared to the control group with the human partner (Levin et al., 2013). Further they found this to be correlated with a significant lower level of ascribing intention to a robot. It seems as if humans tend to experience elevated cognitive conflict when they must interact with robots but at the same time do not trust in their intelligence (Levin et al., 2013). In experimental research, dissonance is often measured by using self-report measures such as questionnaires (Levin et al., 2013) or by response times (Botvinick et al., 2001). In case of the latter, increased response times are commonly interpreted suggesting elevated level of cognitive conflict (Botvinick et al., 2001). This is because the processing of incongruous information seems to take more time than for congruent information (Little et al., 2018).

Now that we have discussed how cognitive conflict arises, how it is measured, how it affects society in a positive or negative manner, and most importantly that it also appears when humans interact with robots, we should ask ourselves: is there a way to improve our experience when collaborating with non-human agents? Can we somehow accomplish that humans experience less cognitive conflict when interacting with robots? With this research project, we try to examine three factors (i.e., agency in decision, outcome value and outcome ownership) in how they shape the level of cognitive conflict when interacting with robots. In the following paragraphs, we will discuss the existing literature on these three factors and relate them to our question of research. We think that it is crucial to understand the interplay of these factors to design better robots and to maximize their positive impact for our society.

2.3 Sense of Agency

Sense of agency refers to the feeling that one's actions and their perceivable outcomes are self-controlled (Gallagher, 2000; Haggard & Eitam, 2015). This feeling is essential for establishing a consistent sense of self (Gentsch & Schütz-Bosbach, 2015), as it enables individuals to recognize themselves as the initiators of their actions, thereby promoting a stable personal identity and a cohesive life narrative (Jakob, 2007). When people feel in control of their actions, they are more likely to experience higher levels of subjective wellbeing and life satisfaction (Kotan, 2010). This is because agency allows individuals to do, what they feel is important, which often leads to outcomes that contribute to their happiness (Kotan, 2010). Moreover, sense of agency plays a significant role in psychological disease. In conditions like schizophrenia, a distorted sense of agency can cause individuals to feel disconnected from their actions (Gentsch & Schütz-Bosbach, 2015), leading to delusions of control. This disconnection can severely impact their ability to function and impair their overall health. Similarly, diminished agency is often seen in depression and anxiety (Mehta et al., 2023), where individuals feel powerless over the circumstances they are living in.

As shown above, sense of agency is a fundamental aspect of human decision-making that influences aspects such as happiness, the formation of self-identity and the manifestation of mental disorders. The influence of self-agency so wide-ranging that we believe it could also play a significant role in individuals' experiences when interacting with robots. But under what conditions do humans experience sense of agency, and how can this phenomenon be effectively elicited and studied in experimental settings? Gallagher (2013) introduces several factors which, in his opinion, create and shape sense of agency in humans. Gallagher (2013) states that "[w]e do not attend to the details of our bodily movements in most actions. We do not stare at our own hands as we decide to use them; we do not look at our feet as we walk …" (p. 11) but we are anyways aware that it is us moving our body. These kinds of non-conscious and automatic motor actions play a crucial role in maintaining a coherent sense of self-agency as a human being. Recognizing the movement of our limbs is one factor that shapes our experience of self-agency, Gallagher (2013) categorizes the factors that influence self-agency into five different groups: motor control processes, the formation of future and present intentions, pre-reflective perceptual monitoring and the retrospective attribution of agency.

The experience of causing a change in the environment by using motor control processes to perform a physical action is very often understood as a constitutional factor for sense of self-agency. In his essay „*Ambiguity in the sense of agency*“, Gallagher (2013) argues that motor control processes are not necessary to experience a sense of agency and introduces the difference between sense of agency and sense of ownership. Gallagher (2013) uses the example of being pushed: in this case one would encounter a high sense of ownership for the bodily movement, but a low level of agency, since there was no prior personal intention to this movement. This leads to the second factor that is essential for self-agency: intentions. Gallagher (2013) argues that there are always ”higher-order cognitive components involving intention formation” (p. 4) before going into a sensory-motor process. He further clusters intentions into different groups that serve different purposes. Future-intentions to form long-time goals, such as planning to buy a new car, present-intentions for short-term intentions, like getting a drink when being thirsty and motor-intentions for rather habitual actions, such as reaching for a glass of water on the table. While present- and future-intentions involve higher levels of agency, motor intentions would usually come along with lower, but still minimal experience of self-agency (Gallagher, 2013). Opposed to the this predominantly conscious process of intention forming, Gallagher (2013) also puts emphasis on subconscious operations influencing sense of agency by introducing the term of pre-reflective perceptual monitoring. In his opinion, when we perform actions, we constantly receive feedback about movements and outcomes from our environment. This feedback is ‘pre-reflective’, because we usually do not actively think about it, but it subconsciously provides us with a sense of self, a sense of agency and ownership as long as the received feedback is in line with our reflected intentions and expectations (Gallagher, 2013). Finally, and in contrast to intentions as well as pre-reflective monitoring, there is the category of the retrospective attribution of agency. While intentions and pre-reflective monitoring always happens previous to an action, movement or event, retrospective attribution of agency allows to update agency experience even after an event or action. Graham and Stephens (1994) are further explaining this phenomenon with the tendency to construct “self-referential narratives” (p. 101) which enables humans to decode their behavior retrospectively. Gallagher (2013) summarizes the influencing factors of self-agency by explaining that “the sense of agency is a matter of degree – it can be enhanced or reduced by physical, social, economic, and cultural factors – sometimes working through our own narrative practices […]” (p. 15).

When relating Gallaghers work to our project, it becomes apparent that including intentions into the experimental setup is an important aspect of investigating sense of agency. Allowing participants to form intentions (e.g., wanting to gain more money) instead of simply following a motor task is crucial for the experience of agency.

Given our understanding of how to effectively elicit sense of agency and its underlying mechanisms in humans, we propose it as a significant factor that influences the processing speed of gains and losses and therefore shapes cognitive dissonance in interactions with robots. A recent study by Giersiepen et al. (2023) explored self-agency in decision-making scenarios with both positive and negative outcomes, revealing distinct brain activity patterns corresponding to low and high agency situations. These patterns suggest that the level of self-agency might alter the cognitive processing of the presented decision-outcome situations, potentially also shaping the cognitive processing speed of the stimuli outcome values.

2.4 Feedback Monitoring and Outcome Evaluation

To determine the outcome value of an action or decision, humans need the ability to monitor feedback. For our study, we propose, following the work by Abubshait et al. (2023), the ‘outcome value’ (i.e., the monetary value a person gains or loses in a trial) as a relevant factor to shape the processing speed of action outcomes and therefore give insights about the level of cognitive conflict in interaction with robots. According to Bandura (1997), feedback monitoring is the process by which humans receive, interpret and subsequently respond to feedback from the environment, which may come from their own actions or from others. The process involves evaluating the feedback using knowledge, beliefs and preconceptions as well as assessing the relevance and impact of the feedback (Bandura, 1997). Feedback often triggers emotions, which are then also referred to as affective feedback, as they are an integral part of the evaluation process and can therefore significantly influence the interpretation of the feedback received (Bandura et al., 2003). In addition to triggering emotions, feedback leads to various outcomes, such as behavioral adjustments, goal setting, motivation and learning. These results help people to prevent or abstain from harm by avoiding punishment or pain based on the feedback in subsequent actions (Bandura, 1997). For example, laughter due to an appropriate and good joke will lead the person to repeatedly show humor, whereas an astonished or even negative expression due to an inappropriate joke will lead to avoidance of such statements. In a learning situation, people use negative feedback to avoid repeating certain behaviors in the future, while positive feedback reinforces actions for the next time (Kaiser et al., 2021). In addition, feedback monitoring helps individuals to adapt to new situations and challenges by refining their understanding and improving their decision-making processes (Bandura, 1997). Previous research has shown that when humans receive feedback in form of monetary gains or losses, they tend to value losses higher than wins (Proudfit, 2015). For that reason, we doubled the value of the win (+4 cents) compared to the loss (-2 cents) in our experimental setup (Proudfit, 2015).

2.5 Outcome Ownership

The last factor we propose to investigate in our research, following the research conducted by Abubshait et al. (2023), which we believe influences the processing speed of rewards and losses and therefore gives insights into the level of cognitive conflict when interacting with robots, is outcome ownership. Outcome ownership refers to the extent to which individuals feel responsible for the results of a decision or action, whereas decision agency relates to the control or authority one has in taking that decision. Previous research found that, when gambling for money, individuals seem to be more invested in a gamble when they are directly affected by the outcome (i.e., if they can keep the money that they win or lose (Hassall et al., 2016). Researchers found that their brain responses are stronger when the outcome, whether it is a 'Win' or 'Lose,' affects themselves compared to when it affects another person (Hassall et al., 2016). This elevated brain reaction suggests that people are more emotionally and cognitively engaged when their own outcomes are at stake (Hassall et al., 2016; Krigolson et al., 2013). Interestingly, the difference in reactions between outcomes for oneself and for another person diminishes when the 'Other' is perceived as a friend rather than a stranger (Leng & Zhou, 2010). This finding, highlighted by Leng & Zhou (2010), suggests that familiarity and intimacy of a relationship play a crucial role in how we process and monitor feedback in long-term social interactions.

Based on this background knowledge Abubshait et al. (2023) conducted a study where participants interacted with a robot in an online gambling scenario where they could either win or lose points for themselves or the robot. They found interesting effects of the outcome ownership on response times, suggesting that it takes participants more time to give away outcomes than keeping them to themselves, suggesting elevated levels of cognitive conflict in these scenarios.

2.6 Online Experiments with PsychoPy

PsychoPy is an open-source software solution that was published in 2002 and allows researchers to set up their own psychological experiments with comparatively little effort (Peirce et al., 2019). Since 2018, PsychoPy was made available for web-based experimental procedures. PsychoJs, which can translate experiments created by the PsychoPy builder into JavaScript (*Open Science Tools Ltd.*, 2019), together with Pavlovia, the hosting service and GitHub version control of PsychoPy, enables online testing of experiments (*Open Science Tools Ltd.*, 2019). The experiment code created in PsychoPy and converted by PsychoJs is uploaded to Pavlovia using an automatic integration and can be piloted and made live from there (*Pavlovia*, 2023). Next to workshops and consultancy units, OpenScienceTools sells so called credits which are necessary to put an experiment live on Pavlovia for testing (*Home — PsychoPy®*, 2024). The platform allows integration with common participant recruitment platforms like Sona and Prolific (*Home — PsychoPy®*, 2024) which were both used to recruit participants for the presented research project.

2.7 Interdisciplinarity

The interdisciplinary nature of this project is evident in its integration of cognitive psychology, neuroscience, sociology, economics, and computer Science. At its core, the project investigates cognitive patterns, specifically the phenomenon of cognitive dissonance, within a psychological settings involving interactions with robots, grounding itself mainly in the field of cognitive psychology. By employing response time measures, which reflect underlying neuronal processing as well as motor control processes and behavioral patterns, the project incorporates methods from behavioral cognitive psychology to gain insights into the cognitive mechanisms at play. Additionally, the project draws on sociology by exploring collaboration scenarios with non-human agents, examining how humans interact and collaborate with robots. These scenarios are further enriched by making use of economic principles, where participants engage in gambling tasks for monetary rewards, leveraging their internal motivation to maximize gains. Finally, the experimental setup required the application of computer science knowledge, particularly in customizing the PsychoPy software with self-written Python code snippets to tailor the experimental conditions. This blend of methods, theories and knowledge from different disciplines not only enhances the robustness of the research but also ensures deep insights into the cognitive phenomena under investigation.

2.8 The Current Study

Based on the study of Abubshait et al. (2023) and Giersiepen et al. (2023) , our aim was to test how sense of agency, outcome value and outcome ownership interact in shaping the cognitive processing speed of participants when interacting with robots. For this purpose, we devised a monetary decision-making task where the participant or the robot chose between two presented stimuli, resulting in either a win or loss of a small amount of money, which was then assigned to either the participant or the robot.

Response times when confirming the outcome owner served as an indirect measure of cognitive dissonance in the decision situation. Our first aim was to confirm previously found effects by Abubshait et al. (2023) regarding the influence of outcome value as well as outcome ownership, which suggested that the outcome ownership, but not the outcome value, significantly shaped the processing speed of rewards and losses. They further found an interaction effect between both investigated variables on response times, suggesting:

1. Faster responses for keeping compared to giving in cases of a positive outcome.
2. Slower responses for keeping compared to giving in cases of a negative outcome.

3. Faster responses for winning compared to losing in cases of keeping an outcome.
4. No significant difference between losing and winning in cases of giving an outcome.

Additionally, we planned to investigate how the newly suggested factor of decision agency interacts and shapes the processing speed of rewards and losses together with the already introduced factors of outcome ownership and outcome value.

2.9 Research Questions and Hypothesis

In the context of the previously discussed theoretical framework, we aimed to investigate how different levels of decision-making agency, different values of action-outcomes as well as different levels of outcome-ownership shape the processing speed of rewards and losses and therefore the level of conflict in interaction situations with robots. We anticipated that low agency in decision together with positive action-outcomes, and outcomes attributed to a non-human agent will elicit decreased response times, indicating reduced levels of cognitive conflict compared to the same situation, but with high decision agency. We believed this would be the case because of a less invested participant when not initially being involved in the decision situation.

III METHODS

3.1 Participants

The study was made available on Sona (*Platform · Sona Systems*, 2002) and Prolific (*Prolific | Quickly Find Research Participants You Can Trust*, 2014). On Sona, participants could choose to either be reimbursed with 9 Euro per hour or participation credits + a bonus of 2.10 Euro from the gambling task. On Prolific, participants got a refund and the bonus of the experiment. To take part in the study, participants needed to confirm that they are between 18 and 45 years old, can comprehend basic English instructions, have normal or corrected-to-normal vision without color vision deficiency, do not intake any medication affecting neural functioning as well as not being affected by any current neurological or psychiatric disorder. Further, all participants provided informed consent before starting with the experiment. In total, 150 participants completed the experiment (61 females, 88 males, 1 other). The data of 9 participants was excluded due to high error rates corresponding to an error in at least 28.92% of trials in the outcome attributions. Thus, the final dataset contained datapoints from 141 participants (57 female, 83 male, 128 right-handed, 1 other-handed) with a mean age of 28.42 years (*SD* = 6.01).

3.2 Experimental Design

3.2.1 High-Level Design

The experimental design let the participant experience different levels of agency through the choice of one of the two presented stimuli which was either performed by themselves (high agency) or the interacting robot Cozmo (low agency). Next, the outcome value (winning or losing a small amount of money) of the subsequent decision was presented to the participant. The affective feedback due to monetary gains or losses held no informative value for learning or adapting behavior successfully. The participants, however, did not know about this and therefore would continue trying to learn a rule by which they think they can maximize the outcome value for themselves.

This was followed by the attribution of the outcome value to either the participant or the robot Cozmo which had to be confirmed by the participant. A schematic overview of the trial sequence can be found in figure 1. The response time of this confirmation served as measure of interest in our online experiment. There were two possible breakout scenarios which lead to a repetition of the subsequent step of the trial: the participant did not confirm correctly, or the response of the participant was too slow. The participants were running through some practice trials, which followed the exact same logic as the experiment trials before starting into the procedure. They must complete four of them successfully without once needing too much time or giving a wrong response. Only then, the trials of the actual experiment commenced.

Figure 1: Screens of the experiment trial procedure

3.2.2 Trial Setup

The condition levels for each trial were hard coded in the setting file of each block. Considering the three independent variables (1) decision agency, (2) outcome value, and (3) outcome ownership with two levels each, one can distinguish between eight experimental condition sets in total:

1. High decision agency + positive outcome + self-owned
2. High decision agency + positive outcome + other owned
3. High decision agency + negative outcome + self-owned
4. High decision agency + negative outcome + other owned
5. Low decision agency + positive outcome + self-owned
6. Low decision agency + positive outcome + other owned
7. Low decision agency + negative outcome + self-owned
8. Low decision agency + negative outcome + other owned

Each testing block held four trials of each condition set, adding up to a total of 32 trials per block, and an overall amount of 384 trials throughout the whole experiment. When considering all twelve blocks, 48 trials of each condition set were presented across all blocks of trials.

The participant would go through the complete set of trials which initially contained the exact same of win and lose trials as well as the exact same outcome attribution towards Cozmo and the participant leading to identical monetary balances for both players. We expected that the participants would quickly realize that there is no way to improve performance by taking “correct” decisions. To account for this, we decided to partially randomize some of the trial outcome values in order to add imbalance to the condition set distribution and thus in the accumulated balance of Cozmo and the participant. In each block, we randomly selected three trials in which we changed the outcome value to the opposite; winning trials ended up being losing trials and vice versa. As follows, approximately 9.38% of trials had a different condition set as expected, which lead to a slightly disbalanced money balance between the participant and Cozmo. Following this logic, the experimental procedure held

- 49 trials of high decision agency + positive outcome + self-owned
- 45 trials of high decision agency + positive outcome + other owned

- 47 trials of high decision agency + negative outcome + self-owned
- 51 trials of high decision agency + negative outcome + other owned
- 50 trials of low decision agency + positive outcome + self-owned
- 46 trials of low decision agency + positive outcome + other owned
- 46 trials of low decision agency + negative outcome + self-owned
- 50 trials of low decision agency + negative outcome + other owned

3.3 Experimental Setup

3.3.1 Implementation

The code used several Excel files which defined the experiment blocks and the trials. All independent variables and resources were hard coded inside those files and the experiment logic randomly selected one of the rows and ran through the corresponding columns to set the variables in the experiment. First, the "experiment blocks overview" file (Figure 2) was used, which defined the different experiment blocks, starting with the practice block, followed by block 1, block 2 and block 3, et cetera.

	A	B	C	D
1	condition_file_name	block_name	row_numbers	
2	ConditionsTest.xlsx	the practice block	0:32	
3	ConditionsBlock1.xlsx	Block 1 of 12	0:32	
4	ConditionsBlock2.xlsx	Block 2 of 12	0:32	
5	ConditionsBlock3.xlsx	Block 3 of 12	0:32	
6	ConditionsBlock4.xlsx	Block 4 of 12	0:32	
7	ConditionsBlock5.xlsx	Block 5 of 12	0:32	
8	ConditionsBlock6.xlsx	Block 6 of 12	0:32	

Figure 2: Setting file – experiment blocks overview file

The file was processed in sequential order (from top to bottom) and contained a column relating to the "detailed experiment block" file for each block where the trials were defined, and which were entered throughout the experiment. Therefore, the code would first use the practice setting file, and all trials defined within it before entering the setting file for the first, second, third block, et cetera.

The experimental logic consisted of two similar main paths: one path for the high agency condition, in which the participant chose one of the colored squares; and another path for the low agency condition, in which Cozmo chose one of the colored squares. This was necessary due to small differences between both scenarios which needed to be represented by a different implementation in PsychoPy Builder of the version v2024.1.4. To take the correct path, the code entered the detailed setting file (Figure 3) and randomly selected one of the hard coded trials. The “chooser” column of the selected trial contained the information which of the paths should be entered for this trial.

	A	B	C	D	E	F	G
1	block	trial	chooser	stim1	stim2	choice_frame_lo	owner	
2	1	29	you	images/stim14_	images/stim13_	-	you	
3	1	19	you	images/stim14_	images/stim13_	-	cozmo	
4	1	15	cozmo	images/stim14_	images/stim13_	-0,3	you	
5	1	31	you	images/stim14_	images/stim13_	-	you	
6	1	5	cozmo	images/stim14_	images/stim13_	0,3	you	
7	1	26	you	images/stim13_	images/stim14_	-	cozmo	
8	1	27	you	images/stim14_	images/stim13_	-	cozmo	
9	1	7	cozmo	images/stim14_	images/stim13_	0,3	you	
10	1	22	you	images/stim13_	images/stim14_	-	you	
11	1	9	cozmo	images/stim14_	images/stim13_	-0,3	cozmo	
12	1	24	you	images/stim13_	images/stim14_	-	you	
13	1	16	cozmo	images/stim13_	images/stim14_	-0,3	you	

Figure 3: Setting file – detail experiment block file

Following the path information in the setting file, the loops around the corresponding paths were enabled or disabled. The procedure, as described in the chapter “3.4 Procedure”, started. The trials of each block further diverged regarding the target stimulus colors (i.e., in each block, the colors stayed the same for the whole testing procedure, only changing after finishing the current block). The location of the colors on the right- or left-hand side of the screen were counterbalanced over all trials of the block. The same is true for the location of choices done by Cozmo. The side of the color that Cozmo was choosing was hard coded in the setting file and equally distributed over all trials of the block. The setting files are, together with the experiment code, available on the GitLab repository of Pavlovia: https://gitlab.pavlovia.org/boslab/daoo_experiment.

The breakout scenarios all made use of the same screens and feedback visuals: red frames were displayed in case of a wrong choice and an information screen was presented with a prompt asking the participant to respond faster. The screens of the breakout scenarios can be found inside the appendix (Appendix A: Breakout screens). The breakout screens were followed by a repetition of the steps that were presented before the error happened, so that the participant must follow the on-screen instructions in order to continue with the experiment.

3.3.2 Colors

The color pairs used for the decision task were previously defined in a similar research project by Giersiepen et al. (2023) by considering the CIELAB color space (Holy, 2024). “To ensure good stimulus discriminability, and to prevent illusory carry-over effects from … previous blocks, all stimuli were created to be maximally perceptually different to each other …” (Giersiepen et al., 2023, p. 4). Each block made use of one set of colors and therefore colors are repeated only within one block for each participant. An overview of the color pairs and their RGB values can be found inside the appendix (Appendix B: Colors practice trials, Appendix C: Colors experiment trials).

3.4 Procedure

Participants started the online experiment with reading through a manual where they were welcomed and informed that they would be playing a decision-making game with Cozmo, a computer-simulated AI robot. They were further instructed that their goal was to earn as much money as possible throughout the experiment. After this introduction, participants were provided with an informed consent and asked to indicate their age, gender, and handedness. Subsequently, the actual procedure started with the first set of trials for practice purposes and continued with twelve testing blocks, each consisting of 32 trials (N = 384 trials).

The trials started with a fixation cross displayed for one second in the center of the screen. Next, a human or robot avatar together with the instruction “You choose” or “Cozmo chooses” was displayed, informing participants about the choice condition of the current trial see “Appendix D: Avatars”. This was followed by the display of two, colored squares adjacent to the left and right of a fixation cross in the screen center for one second. In the high agency condition (where the participant could choose freely), the selection of a colored square could be indicated by using the letter “g” of the keyboard corresponding to the left square and the “k” corresponding to the right square. After the participant's input, which was done with the left or right index finger, the choice was highlighted with a white frame around the stimulus which did stay on the screen for 0.5 seconds. In the low agency condition, the choice of Cozmo did appear as a black frame, together with the text “Please confirm:” on the screen. The participant again used “g” or “k” to confirm or decline Cozmo’s choice; in case of a positive confirmation the black frame did turn white, in case of an incorrect confirmation the wrong choice was highlighted with a red frame and the trial jumped back to the confirmation request. The red frame as visual information was also used for choices or confirmations that took more than one second and therefore were too slow. The participant was informed about the too slow response with the text “respond faster”, which did stay on the screen for one second.

In case of a successful choice sequence, the trial was continued by displaying the outcome value of the chosen stimuli, indicated by a gain of 4ct or a loss of 2ct for one second on the screen. This was followed by an instruction to either attribute the outcome to Cozmo or to oneself by pressing the g-key to “give” the outcome to Cozmo or the k-key “keep” the outcome to oneself, respectively. When participants confirmed the outcome assignment within one second a white frame around the word “give” or “keep” confirmed their response. In case of a wrong or too slow response a red colored frame was presented. Outcome attributions were repeated until participants correctly assigned the outcome to the correct player.

In case of a correct outcome attribution, participants were informed about their and Cozmo’s total current balance as of now by showing the two avatars of the participant and Cozmo on the screen as well as the accumulated gained monetary value underneath both avatars for one second. After that, the next trial was initiated with the fixation cross as detailed above.

In the beginning and end of each block, info screens containing information about the consecutive number of the current block were displayed. Participant could take a break of up to 3 minutes between the blocks or use the “space”-key to immediately continue with the experiment. Monetary balances accumulated throughout one block and were reset when entering a new block.

After participants completed the twelfth block, they were informed of their total balance, which would be paid out to them after the experiment. They could use the “space” key or wait for three minutes to enter another PsychoPy survey module where he or she needed to fill out two additional short questionnaires, namely the Center for Epidemiologic Studies Depression Scale Revised with the short name CESD-R (Eaton et al., 2004) and the Internale-Externale-Kontrollüberzeugung-4 with the abbreviation IE-4 (Kovaleva et al., 2014). Participants were thanked for participation as well as informed about possible contacts for feedback. The CESD-R questionnaire screens for depression and depressive disorders, the IE-4 questionnaire is assessing the psychological characteristics of control conviction. The data from these questionnaires is not the focus of the current research project but will be addressed in a subsequent follow-up project. An overview of the time settings for the experimental procedure can be found inside the appendix (Appendix E: Procedure time settings).

The complete code can be viewed and cloned on the GitLab repository of Pavlovia with the address: https://gitlab.pavlovia.org/boslab/daoo_experiment.

IV DATA ANALYSIS

4.1 Data Preprocessing

4.1.1 Questionnaire Data

The preprocessing logic for the questionnaire data started by checking each data file one by one, ensuring that all needed columns were available. In case of a missing column (e.g., because some of the paths, like the “respond faster”, weren’t accessed in the experiment) the missing column was created and filled with NA values. Next, the data was split in several subsets. This was necessary as PsychoPy did not interpret the experimental paths correctly. The created subsets are used to join the data of one trial into one row. Further, columns were renamed so that researchers could understand which column represented which question. The same is true for the values within the columns, which represented the answer options for participants; the standard values from PsychoPy were mapped to interpretable values. The data dictionary of the preprocessed dataset can be found in the appendix (Appendix F: Data dictionary – questionnaire data).

4.1.2 Response Time Data

In the beginning, the preprocessing of the response time data followed the same logic as the one from the questionnaire data. Next, some of the columns were combined into single ones since they contained similar information. For instance, the response times for confirming the outcome attribution in the You-Choose and Cozmo-Chooses path were initially stored in two separate columns. By means of preprocessing, these are merged into a single column. This was followed by the correction of the outcome values. The setting files contained the information which of the trials were randomly selected for changing their outcome value into the opposite, but the outcome value column was not yet adjusted. To ease up analysis, the outcome value as well as the condition identifier were therefore corrected by changing “win“ into “lose“ and “lose“ into “win“ in case of a chance trial. Further, the offset and onset of the “too slow“ or “wrong response“ path were meant to give information regarding if those paths were accessed or not. Therefore, their content was changed into Boolean values (“true“ / “false“) instead of holding the on and offset times; when no time was available, the path was not accessed and when a time was available, the path was entered in the experiment. This allowed to determine the error rates of participants. Some standard preprocessing steps, such as renaming columns and parsing of date and time, were performed. Before the final columns of interest were selected from the dataset an index was created, indicating the order of the trials within blocks. The data dictionary of the preprocessed dataset can be found in the appendix (Appendix G: Data dictionary – response time data). The preprocessing let us end with short names for the independent variables. CHOOSER represents the factor decision agency and holds the possible features ‘participant’ and ‘cozmo’, VALUE represents the outcome value with the possible features of ‘win’ and ‘lose’ and OWNER the outcome owner with the features of ‘keep’ and ‘give’.

4.1.3 Error Rates

Participants could make several errors during the experiment: they could choose a color too slowly; they could fail to confirm the color choice of Cozmo or could be too slow in the color choice confirmation, or they could fail to confirm the outcome attribution or be too slow when confirming the outcome attribution. To decide if datasets from participants must be excluded, error rates were calculated. On average, participants made errors in 3.44% (*SD* = 2.90) of the trials, when the color had to be chosen, and in 10.26% (*SD* = 7.47%) of the trials when the outcome attribution had to be confirmed by the participant. Based on these findings, we decided to exclude participants who made too many errors in the outcome attribution and set the limits at +/- 2.5 SD of the outcome attribution error rate. Additionally, we excluded the erroneous trials of the remaining participants in which mistakes in the outcome attribution happened as well as trials in which the timing of displayed elements was off. On average, 90% of trials remained in the dataset for all further analyses. The average amount of trials per participant was 346.90 (min: 275, max: 381).

4.2 Descriptive and Statistical Analysis

4.2.1 Re-Examination of Previous Results

One of the goals of this project was to re-examine previously found effects. While Abubshait et al. (2023) used a repeated measures ANOVA with follow-up t-tests to analyze the effects inside their data, we decided, due to its higher sensitivity, to employ hierarchical linear models on our data. The paired interaction (VALUE x OWNER) was tested for the direction of effects by fixing the characteristic of one of the variables and employing a hierarchical linear model on the other remaining variable.

4.2.2 Including the New Factor: Decision Agency

An important additional goal in the current study was to examine especially the effects of decision agency, in combination with outcome attribution and outcome value on response times. We again used the results of the hierarchical linear model, also accounting for the interaction effects between all three variables. Further, the paired interactions (CHOOSER x OWNER and CHOOSER x VALUE) were tested for their directions by fixing the value of one of the variables and again employing a hierarchical linear model on the one remaining factor.

4.2.3 Win-Stay, Lose-Shift Behavior

A common pattern observed in decision-making experiments is the "win-stay, lose-shift behavior" (Cassotti et al., 2011; Worthy et al., 2013). The theory posits that participants are likely to base their decisions on their most recent experiences, maintaining their choice after a win and changing it after a loss (Worthy et al., 2013). In our experiment, it would be intriguing to determine if this behavior was evident when Cozmo made the choice, or if participants only regarded their own decisions as informative. To examine our dataset for this trend, only those trials, in which the participant was the outcome owner were considered as initial trials. For the subsequent trial, only those trials were included, in which the participant was the chooser. After merging the initial and subsequent trial into one row, ensuring that only directly consecutive trials were combined, we could compare the two decisions to determine if a participant chose the same or a different color. This dataset allowed us to calculate the ratios for each participant, indicating how often they selected the same or a different color, while also considering the initial chooser and outcome value. To assess significant trends in the data, we used a hierarchical linear model on both the trials in which participants stayed with the previous choice and on the trials in which participants shifted their choice.

4.2.4 The Influence of the Monetary Balance on Response Times

We further explored whether the current monetary balance of a participant did influence the degree of cognitive conflict when interacting with robots. While piloting the experiment, the idea came up that participants might found it easier to confirm the outcome attribution towards the robot Cozmo if they already had a higher balance than the robot. For this reason, we calculated the current balance of the participant and the robot at each point in the experiment and compared the participants balance with Cozmo's balance to determine in which trial the participant owned more than Cozmo. This dichotomous information was then used as an explanatory feature in a hierarchical linear model that we employed to compare the response times of trials in which the outcome was attributed to the robot.

4.2.5 Response Times Over the Course of the Experiment

To understand potential learning effects or fatigue throughout the testing phase we analyzed the response time data of all participants over the course of the experiment. For this, it is important to note that trials were grouped by their index within the experiment, which is not necessarily a time related information. For example, the trial with the index 100 of each participant could be seen at the second 500, 600 or even 660 after starting the experiment. It depended on how many mistakes a participant made and how long it took them to complete the trials before the hundredth trial. Nevertheless, we assumed that the index, as an approximation of a timeline, provided information about how the performance of participants developed over time.

4.2.6 The Influence of Handedness on Response Times

To consider how the handedness of participants did influence our results, we analyzed how handedness related to response times. We hypothesized that right-handed people might have been faster in selecting the stimuli on the right-hand side, confirm Cozmo's selection of the right stimuli, as well as confirm a “keep“, because all of them have been selected with the right index finger. The same but vice versa would be true for left-handed people, using the “g“ with the left index finger to confirm and select items at the left side of the experiment screen or to “give“ outcomes to Cozmo. To do so, response times of each participant in the outcome attribution were averaged over all trials for both keys, “k“ and “g“. Then, participants were grouped by their handedness and their individual average response times again averaged over all participants. The one participant who chose "other" as his or her handedness was discarded for this analysis. To assess significant trends in the data, we utilized a hierarchical linear model that also accounted for interaction effects.

4.2.7 The Influence of Handedness on Choice Location

A similar assumption was made for the correlation between handedness and the location of the choice. We hypothesized that right-handers would prefer the stimuli on the right-hand side of the screen to the one on the left and would therefore choose it more often. Respectively, we expected the same to be true for left-handed participants and the left square of the screen. To analyze this, we only looked at the samples in which the participant was the one choosing the color and calculated the ratio of left and right choices per participant. The ratios were then averaged and again grouped by handedness across all participants. Again, the one participant that indicated "other" as his or her handedness was discarded for this analysis. To check the descriptive statistics for significant differences, we applied a chi squared test.

4.2.8 Repetition-Driven Response Times

To gain insight into how repetition influenced response times in our experiment, we analyzed how response times behaved after twice using the same key. In other words, when a participant chose or confirmed the color choice with “g“ and then had to “give“ the outcome, he or she could potentially be faster in response because of the repetition (Bentin & McCarthy, 1994). To analyze this, an identifier was created for all the trials in which the choice location and the owner confirmation was done with the same key. This identifier was then used to group the trials into the identical and non-identical group for each participant. Then, response times were averaged first for each participant and in a second step over all participants. To assess significant trends in the data, we utilized a hierarchical linear model.

4.3 Analysis Software and Packages

The data analysis is published on GitHub (*GitHub*, 2008) in a publicly available repository: https://github.com/anjaghub/master_thesis and was logically separated in several files. The files were written in Python version 3.9.19. The preprocessing of data was split into two files, one of them containing the preparation of the data from the questionnaire in the beginning and end of the experiment ("survey_data.py"), and another file containing the preparation of the response time data ("data_prep.py"). For the descriptive and statistical analysis, a Jupiter notebook ("analysis.ipynb") was used. The "requirements.txt" file gives insights about all employed python libraries and the "readme.md" introduces users to the purpose and usage of the repository. Further, a pre-analysis of the data from the pilot study was done inside the "data_check.py" file to ensure counterbalancing of all variables and availability of required datapoints for the analysis. A separate function that was implemented to sort out trials in which the timing of elements was off is available in the "timing_check_function.py" file.

For the hierarchical linear models we used the mixedlm method from the statsmodel package (version 0.14.1; Seabold & Perktold, 2010) following the approach described by Lindstrom and Bates (1988). To perform the chi squared test we made use of the chi2_contingency method from the SciPy package of the version 1.14.0 (*Scipy.Stats.Chisquare*, 2024).

V RESULTS

5.1 Confirming Previously Found Effects

To replicate and maybe confirm the results previously found by Abubshait et al. (2023), we checked our data regarding the main effects of OWNER and VALUE as well as their interaction effect. The distribution of response times of all participants and all trials grouped by the independent variables OWNER and VALUE (Figure 4 and 5) gave some first insights into these questions. When observing the response time distributions of the outcome ownership (Figure 4), one could see that the “give“-curve was slightly shifted to the right and the “keep“-curve had a higher frequency characteristic on the left side of the distribution curve.

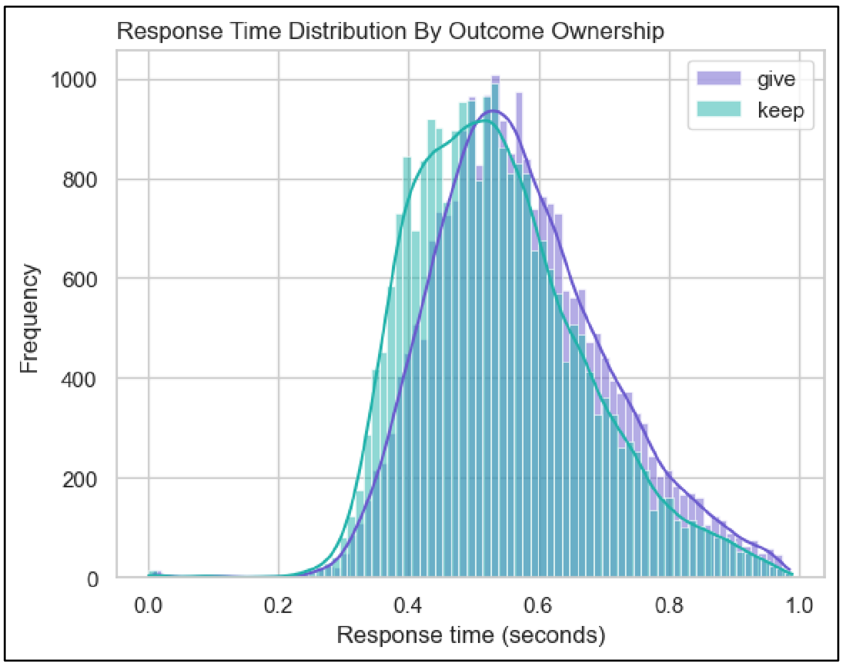


Figure 4: Response time distribution comparison of outcome ownership

When analyzing the response time frequency distribution of the outcome value groups (Figure 5), one could observe a similar but weaker shift between the curves. Here, the “winning“-curve again was slightly left shifted compared to the “loosing“-curve.

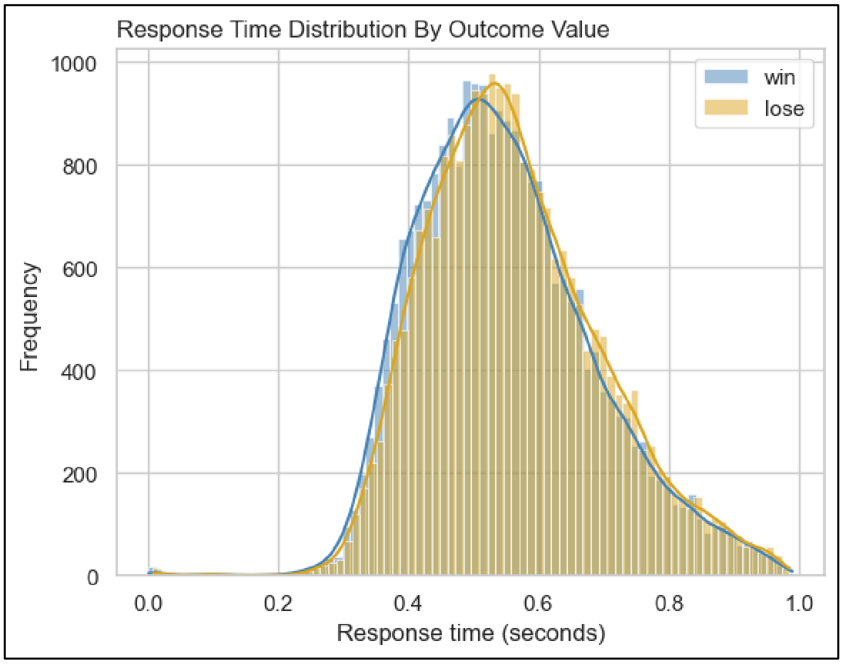


Figure 5: Response time distribution comparison of outcome values

To test the observed trends for significance, we analyzed the results of the hierarchical linear model. The information on the applied model (Table 1) suggested that the fitting process was successful, and the high log-likelihood indicated a good fit of the model.

Table 1: Mixed linear model information – effect of chooser, value and owner on response time

model:	mixedlm	dependent variable:	owner_confirm_rt
no. observations:	48906	method:	REML
no. groups:	141	scale:	0.0137
min group size:	275	log-likelihood:	35068.49
max group size:	381	converged:	yes
mean group size:	346.90		

As baseline condition set Cozmo as chooser and owner, as well as a negative outcome value was set. The baseline response time for this condition set was 0.59 seconds. The model results (Table 2) suggested that:

1. When the participant was the outcome owner, the response time decreased on average by 0.03 seconds compared to the baseline. This effect was significant with $p < .001$ and confirmed the found effect by Abubshait et al. (2023).
2. There was no main effect of VALUE on response time ($p > .05$). This confirmed the observation by Abubshait et al. (2023).
3. There was a significant interaction effect between OWNER and VALUE with $p < .001$ which confirmed the found interaction effect by Abubshait et al. (2023).

Table 2: Mixed linear model output (1) – effect of value and owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.59	0.01	109.01	0.00	0.58	0.60
owner [keep]	-0.03	0.00	-11.78	0.00	-0.03	-0.02
value [win]	-0.00	0.00	-1.11	0.27	-0.01	0.00
owner[keep]: value[win]	-0.03	0.00	-8.19	0.00	-0.03	-0.02
Group Var	0.00	0.00				

Next, the absolute differences of the response time averages were checked (Table 3). When comparing them for keep-win to keep-loss, one found a difference of 0.03 seconds in response times, while the difference between give-win and give-loss is only 0.01 seconds. Further a larger difference between win-give and win-keep of 0.06 seconds compared to 0.02 seconds between loss-give and loss-keep can be observed.

Table 3: Response times by outcome attribution and outcome value

outcome attribution	*value*	*average response time (seconds)*	*SD of response time*
give	lose	0.57	0.14
give	win	0.58	0.13
keep	lose	0.55	0.13
keep	win	0.52	0.13

To further understand whether our data replicated the direction of the interaction effects, the two-way interactions with one fixed variable were analyzed:

- In cases where the participant kept the outcome to herself, the response time significantly decreased by on average 0.03 seconds when the outcome was positive (Table 4). This confirmed the found effect by Abubshait et al. (2023).
- In cases where the participant gave the outcome to Cozmo, the response time significantly increased by on average 0.01 seconds when the outcome was positive (Table 5). This difference was not found by Abubshait et al. (2023).

Table 4: Mixed linear model output – participant kept outcome; effect of value on response time

	coef.	*Std. err.*	*Z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.56	0.01	105.03	0.00	0.54	0.57
value[win]	-0.03	0.00	-21.76	0.00	-0.04	-0.03
group Var	0.00	0.00				

Table 5: Mixed linear model output – participant gave outcome; effect of value on response time

	coef.	*Std. err.*	*Z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.57	0.01	102.89	0.00	0.56	0.58
value[win]	0.01	0.00	5.72	0.00	0.01	0.01
group Var	0.00	0.00				

- In cases where the outcome was positive, the response time significantly decreased by on average 0.06 seconds when the participant kept the outcome to herself (Table 6). This confirmed the found effect by Abubshait et al. (2023).
- In cases where the outcome was negative, the response time significantly decreased by on average 0.02 seconds when the participant kept the outcome to herself (Table 7). The effect found by Abubshait et al. (2023) went into the other direction.

Table 6: Mixed linear model output – positive outcome; effect of owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	106.44	0.00	0.57	0.59
owner[keep]	-0.06	0.00	-37.68	0.00	-0.06	-0.05
group Var	0.00	0.00				

Table 7: Mixed linear model output – negative outcome; effect of owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.57	0.01	109.17	0.00	0.56	0.58
owner[keep]	-0.02	0.00	-9.73	0.00	-0.02	-0.01
group Var	0.00	0.00				

All these effects were significant with $p < .001$. The detailed model information can be found in Appendix H.

5.2 Decision Agency Influences Response Times

The violin plots (Figure 6) showed response times grouped over all sets of conditions, combined into one identifier as chooser_owner_value, giving first insights into the overall response times when combining all three independent variables. At a first glance, the distributions appeared very similar, only the “participant_keep_win“ distribution showed a defined frequency characteristic at the lower end of the distribution. It is important to note that the violin plot provided only an initial overview of the data distribution across all participants, but it did not account for individual differences between participants. These differences were considered by using the hierarchical linear models.

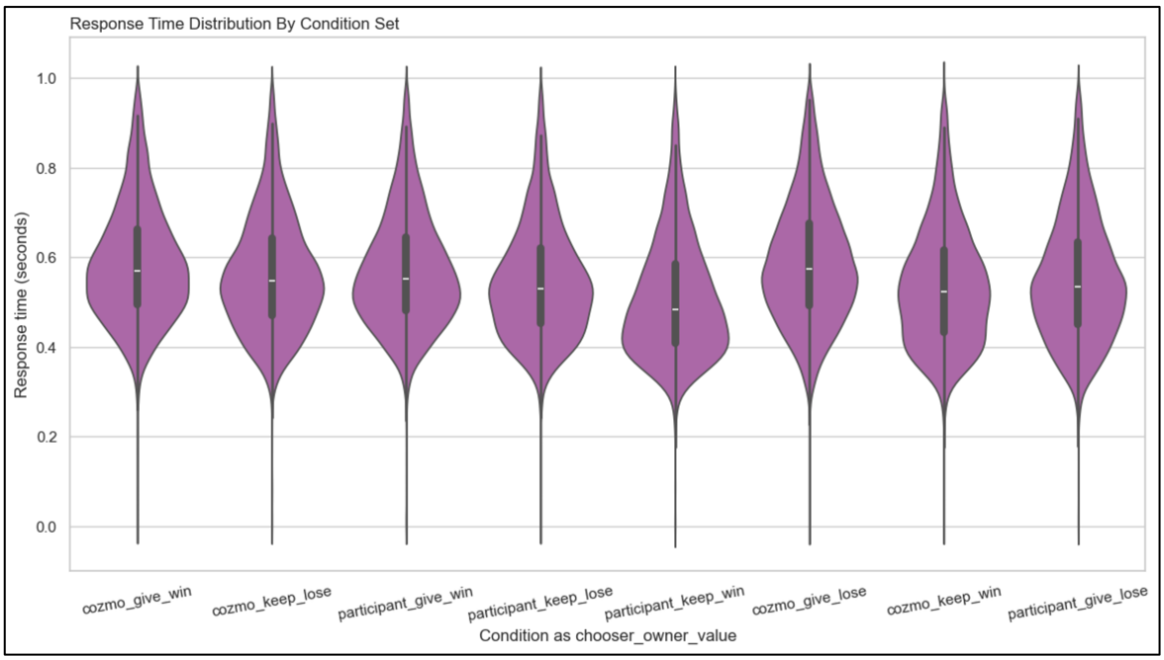


Figure 6: Response time distribution of all condition combinations

The hierarchical linear model allowed for interpretation of significant differences and their directions. As baseline condition set Cozmo as chooser and owner, as well as a negative outcome value was set. The baseline response time for this condition set was 0.59 seconds. The model results (Table 8) suggested that:

1. When the participant was the chooser, the response time decreased on average by 0.04 seconds compared to the baseline condition set.
2. When the participant is the chooser and outcome owner and the outcome value was positive, the response time again decreased on average by 0.03 seconds compared to the baseline condition set.

All those effects were significant with $p < .001$ and the narrow confidence intervals suggested high reliability in the results.

Table 8: Mixed linear model output (2) – effect of chooser, value and owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.59	0.01	109.01	0.00	0.58	0.60
chooser [participant]	-0.04	0.00	-17.97	0.00	-0.04	-0.03
chooser[participant]: owner[keep]	0.02	0.00	6.75	0.00	0.01	0.03
chooser[participant]: value[win]	0.02	0.00	7.11	0.00	0.02	0.03
chooser[participant]: owner[keep]: value[win]	-0.03	0.00	-7.69	0.00	-0.04	-0.02
Group Var	0.00	0.00				

The two-way interactions with one fixed variable value (Table 9 – 16) suggest that:

- In cases where the participant kept the outcome to herself, the response time significantly decreased by on average 0.02 seconds when she chose herself (Table 9).
- In cases where the participant gave the outcome to Cozmo, the response time significantly decreased by on average 0.03 seconds when the participant chose herself (Table 10).

- In cases where the outcome was positive, the response time significantly decreased by on average 0.02 seconds when the participant chose herself (Table 11).
- In cases where the outcome was negative, the response time significantly decreased by on average 0.03 seconds when the participant chose herself (Table 12).

Table 9: Mixed linear model output – participant kept outcome; effect of chooser on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.55	0.01	103.98	0.00	0.54	0.56
chooser[participant]	-0.02	0.00	-15.03	0.00	-0.03	-0.02
group Var	0.00	0.00				

Table 10: Mixed linear model output – participant gave outcome; effect of chooser on response time

	coef.	*Std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	106.03	0.00	0.58	0.60
chooser[participant]	-0.03	0.00	-18.14	0.00	-0.03	-0.02
group Var	0.00	0.00				

Table 11: Mixed linear model output – positive outcome; effect of chooser on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.56	0.01	103.26	0.00	0.55	0.57
chooser[participant]	-0.02	0.00	-14.78	0.00	-0.03	-0.02
group Var	0.00	0.00				

Table 12: Mixed linear model output – negative outcome; effect of chooser on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	110.39	0.00	0.57	0.59
chooser[participant]	-0.03	0.00	-18.18	0.00	-0.03	-0.03
group Var	0.00	0.00				

- In cases where the participant chose herself, the response time significantly decreases by on average 0.03 seconds when she kept the outcome to herself (Table 13).
- In cases where the participant chose herself, the response time significantly decreases by on average 0.01 seconds when the value was positive (Table 14).
- In cases where the robot Cozmo chose, the response time significantly decreased by on average 0.04 seconds, when the participant kept the outcome to herself (Table 15).
- In cases where the robot Cozmo chose, the response time significantly decreased by on average 0.02 seconds, when the outcome was positive (Table 16).

Table 13: Mixed linear model output – participant chose; effect of owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.56	0.01	101.49	0.00	0.55	0.57
owner[keep]	-0.03	0.00	-22.66	0.00	-0.04	-0.03
group Var	0.00	0.00				

Table 14: Mixed linear model output – participant chose; effect of value on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.55	0.01	99.52	0.00	0.54	0.56
value[keep]	-0.01	0.00	-7.66	0.00	-0.01	-0.01
group Var	0.00	0.00				

Table 15: Mixed linear model output – Cozmo chose; effect of owner on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	112.82	0.00	0.58	0.60
owner[keep]	-0.04	0.00	-25.23	0.00	-0.04	-0.04
group Var	0.00	0.00				

Table 16: Mixed linear model output – Cozmo chose; effect of value on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	110.75	0.00	0.57	0.59
value[win]	-0.02	0.00	-10.79	0.00	-0.02	-0.01
group Var	0.00	0.00				

All these effects were significant with $p < .001$. The detailed model information can be found in Appendix H.

5.3 Control Analyses

5.3.1 Win-Stay, Lose-Shift Behavior Even When Robot Chooses

When observing the descriptive results of the win-stay lose-shift behavior (Table 17), one could see that in two condition sets (Cozmo-win-true and participant-win-true) participants preferred the same color choice as opposed to switching their color choice.

Table 17: Win-stay, lose-shift behavior

chooser	*value*	*stay with choice*	*frequency*
Cozmo	win	true	14.45%
participant	win	true	15.32%
Cozmo	lose	true	11.78%
participant	lose	true	12.16%
Cozmo	win	false	12.38%
participant	win	false	11.69%
Cozmo	lose	false	12.30%
participant	lose	false	11.18%

The results of the mixed linear model (Table 18) provided support for the descriptive trend in our data. The model details can be found in Appendix I. These results suggested that for the staying behavior (i.e., sticking to the same choice as previously) only VALUE ($p < .01$) but not CHOOSER ($p > .05$) had a significant influence on subsequent item choices.

Table 18: Mixed linear model output – effect of chooser and value on stay ratio

	coef.	*std. err.*	*z*	*p*>|*z*|	*[0.025*	*0.0975]*
intercept	0.12	0.00	35.49	0.00	0.11	0.12
chooser[participant]	0.00	0.00	0.82	0.41	-0.01	0.01
value[win]	0.03	0.00	5.74	0.00	0.02	0.04
chooser[participant: value[win]	0.01	0.01	0.74	0.46	-0.01	0.02
group Var	0.00	0.00				

The results of the mixed linear model for the shift behavior (Table 19); model details can be found in Appendix J; suggested that only CHOOSER ($p < .05$) but not VALUE ($p > .05$) had a significant influence. Participants were less likely to shift their choice if they were the chooser in the first place, no matter what outcome value their choice had.

Table 19: Mixed linear model output – effect of chooser and value on shift ratio

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.12	0.00	38.16	0.00	0.12	0.13
chooser[participant]	-0.01	0.01	-2.46	0.01	-0.02	-0.00
value[win]	-0.00	0.01	-0.17	0.86	-0.01	0.01
chooser[participant]: value[win]	-0.01	0.01	-0.90	0.37	-0.02	0.01
group Var	0.00	0.00				

5.3.2 Current Monetary Balance Influences Response Times

To answer whether the current balance influenced response times we applied a hierarchical linear model. The modelling results (Table 20); model details can be found in Appendix K; suggested that the balance had a significant influence on the response time of the participant. Participants showed an average response time of 0.58 seconds when their current balance was less or equal to the one of Cozmo (baseline condition). If participants balance exceeded the one of Cozmo response times decreased by an average of 0.01 seconds with $p < .001$.

Table 20: Mixed linear model output – effect of monetary balance difference on response time

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	104.02	0.00	0.57	0.59
participant_owns_more _than_cozmo[true]	-0.01	0.00	-4.87	0.00	-0.01	-0.00
group Var	0.00	0.00				

5.3.3 Faster Response Times at the End of the Experiment

When observing the response times over the duration of the experiment (Figure 7), one could see that participants' average response time decreased with task progression. Moreover, one could observe that the distribution gets a bit narrower towards the end of the trial sequence, suggesting that there was less variance between response times towards the end of the experiment.

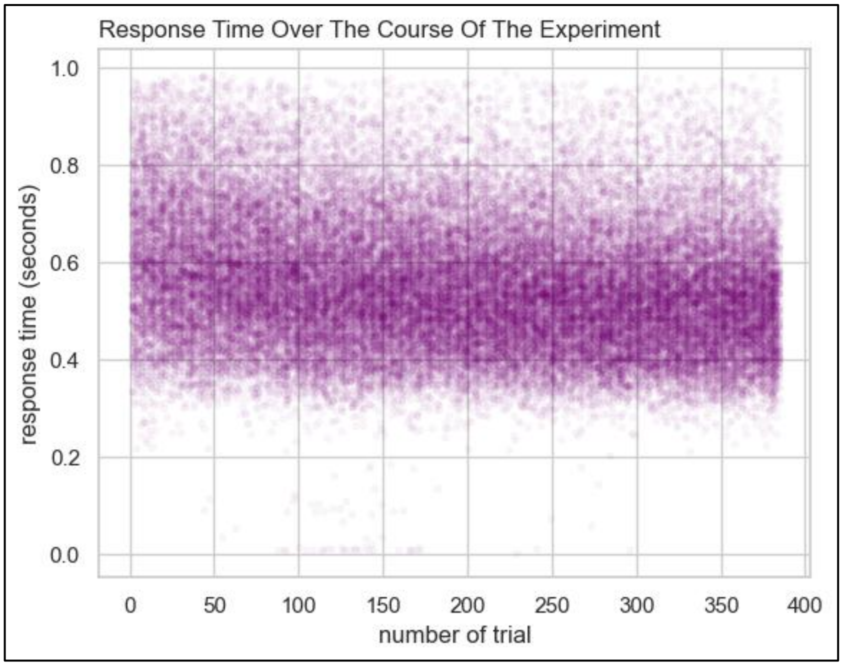


Figure 7: Response times over the course of the experiment

5.3.4 No Influence of Handedness on Response Times

When examining the averaged response times by handedness and location of the outcome assignment key (Table 21), we could see that both right- and left-handed participants were overall faster when they confirmed with the right index finger (using the “k“-key to keep outcomes for themselves).

Table 21: Relation between handedness and response time

handedness	*number participants*	*outcome attribution key*	*average response time (seconds)*	*average SD of response time*
right	128	right (k)	0.54	0.12
right	128	left (g)	0.57	0.12
left	12	right (k)	0.55	0.12
left	12	left (g)	0.59	0.12

The results of the mixed linear model (Appendix L) suggested that there is neither a significant influence of handedness ($p > .05$) nor a significant interaction effect between handedness and the outcome attribution key ($p > .05$) on response time. Solely the outcome attribution key as single feature seemed to influence the response times significantly, with keeping ("k"-key) being faster than giving ("g"-key). Therefore, we could assume that response times did not depend on the handedness of participants which is consistent with our research hypothesis. It is important to note that we had far fewer left-handers than right-handers and that it would be important to re-examine this finding with a larger sample of left-handers.

5.3.5 Influence of Handedness on Choice Location

The results regarding the participants' choices and their handedness (Table 22) indicated that right-handed participants chose more often the right item and left-handed participants more often the left.

Table 22: Relation between handedness and choice location

handedness	*number participants*	*choice location (key)*	*frequency*
right	128	right (k)	51.77%
right	128	left (g)	48.23%
left	12	right (k)	48.40%
left	12	left (g)	51.59%

The Chi-squared test suggests a significant relation between handedness and choice location ($\chi^2(1) = 7.03$, $p = .008$). This means that, in combination with the proportion analysis (Table 21), both left and right handers preferred to choose items with their dominant hand.

5.3.6 No Influence of Repetition on Response Times

When looking at the Table 23 summarizing response times for the repeated key usage (same key for choice and confirmation of outcome attribution), one could see that the average response times in both states were identical up to two decimals. The results of the mixed linear model (Appendix M) suggested that repetition in key usage ($p > .05$) had no significant influence on response times.

Table 23: Repetition driven response times

repeated key usage	*average response time (seconds)*
true	0.58
false	0.58

VI DISCUSSION

The interaction between humans and robots is an increasingly significant area of study, given the inevitable integration of robots into our daily lives. Especially because a lot of people experience a general aversion towards the interaction with robots (Broadbent, 2017), it is crucial to understand how to optimize their collaboration in various scenarios. One aspect that may influence the amount of conflict when interacting with robots is the degree to which a task and its components are outsourced to the robot. Cognitive conflict, defined as the mental discomfort when experiencing conflicting cognitions, is commonly inferred indirectly through response times. Accordingly, increased response times are usually thought to indicate higher levels of conflict (Botvinick et al., 2001). This research project investigated on how decision agency, outcome value and outcome ownership shape the processing speed of rewards and losses, indicating varying levels of cognitive conflict when interacting with robots.

Previous research by Abubshait et al. (2023) has already explored the effects of outcome ownership and outcome value on cognitive dissonance. The findings of this study showed that participants were responding faster when they were keeping an outcome as opposed to when they were giving it away, and that they were faster when confirming a positive outcome as opposed of a negative one (Abubshait et al., 2023). The shorter response times that occurred when participants were processing rewards and losses were interpreted to originate from lower levels of cognitive conflict in these conditions. Further, it was found that it takes participants especially long to confirm the attribution of a positive outcome towards the robot that they interacted with, which was confirmed by a significant interaction effect between both variables on response times (Abubshait et al., 2023).

In the present study, decision agency as a novel variable of interest was introduced. We investigated how low and high decision-agency in combination with positive and negative outcome values as well as different outcome ownerships (self-owned vs. owned by the robot) influenced cognitive conflict during the interaction with robots. We developed an online experimental framework using PsychoPy, in which the participant and the robot chose between two presented stimuli, resulting in either a win or loss of a small amount of money, which was then assigned to either the participant or the robot. The participants' response times for confirming the trial outcome served as an indirect measure of cognitive dissonance for the specific condition set of the trial.

Our findings partly corroborate the results from previous research by Abubshait et al. (2023). We could confirm the suggested main effect of outcome ownership on response time, suggesting an increased processing time and therefore higher level of cognitive conflict in cases where the outcome is given to the robot compared to when they are kept to oneself. Moreover, alike Abubshait et al. (2023), we found an interaction effect of both suggested variables but no main effect from the outcome value on the response time. When looking into the two-way interactions, we could partly confirm the suggested directions by Abubshait et al. (2023). When only looking at win trials, we could confirm that participants are faster in keeping an outcome than in giving it away. Furthermore, when only looking at keep trials, we could confirm that participants were faster confirming positive outcomes than negative ones. Additionally, we found another significant effect that was not present in the study by Abubshait et al. (2023): when only considering give trials, our data suggested significant faster confirmations for negative outcomes than for positive ones. This association suggests that participants preferred giving losses instead of wins to the robot Cozmo. In contradiction to the findings by Abubshait et al. (2023), we found that when only considering lose trials participants still preferred keeping an outcome to themselves over giving it away. Summarizing all these results, this could be interpreted as possessive behavior of humans when interacting with robots, in which humans prefer to control all components of the decision-action situation themselves. Our data seem to suggest that participants tend to prefer keeping outcomes to themselves, especially positive ones, even if the value is negative. Self-owning was always preferred over giving outcomes away. Furthermore, when participants had to give away outcomes, they preferred to give the robot a negative outcome rather than a positive one, revealing a somewhat malicious or egoistic behavior. Comparing the results of our study with those of Abubshait et al. (2023), it can be concluded that it seems very likely that outcome value and outcome ownership influence the processing speed of rewards and losses. Future research should again focus on the direction of the interaction effect to clarify the inconsistencies comparing our results to the ones from previous research.

Next to replicating the results from Abubshait et al. (2023), we introduced and analyzed how the third factor, namely decision agency, shaped the researched phenomena. We found that next to the outcome owner, also the chooser had a significant influence on response time. Similar as previously reported for the factor outcome ownership, the model suggested that participants showed faster confirmations when taking the initial decision themselves. Next to the main effects of the single factors, the model revealed significant interaction effects between all three variables. The chooser together with the outcome owner, the chooser together with the outcome value, as well as all three factors together showed significant interactions influencing the response times of the participants. When interpreting the output of the hierarchical model for the directed two-way interactions, one could understand that when looking into the two cases of Cozmo as chooser and the participant as chooser, participants again showed decreased response times for positive outcomes. Moreover, for both cases, Cozmo as chooser or the participant as chooser, participants showed decreased response times for keeping an outcome compared to giving an outcome. And finally, no matter if looking only into keep trials, give trials, win trials or lose trials participants showed decreased response times when the trial was started with the initial choice by themselves. This trend could be interpreted as a more invested and less conflicted participant when the trial with the initial choice was started by participants themselves.

When looking into the three-way interaction, one can observe that participants confirmed fastest when they were the chooser and the owner of a positive outcome. Overall, these trends and interactions indicated a tendency to seek control in interaction scenarios with robots. Be it in the decision-making phase or in the outcome attribution phase of an action, preferring positive outcomes for ourselves and negatives for the interacting robot. Comparing the directed effect of agency on response time when only looking into trials with an outcome attributed to the robot to our initial hypothesis (see chapter 2.9) that low agency in decision, positive action-outcomes, and outcomes attributed to a non-human agent will elicit decreased response times, one notices that we could not proof or initial hypothesis. Decision agency did have an influence on the response times, but in a different direction than expected. Low agency in decision seemed to not let our participants being less invested and therefore experience lower cognitive conflict. Instead, they seemed to prefer high control also in situations where the robot owns the outcome by having high decision agency.

To ensure the validity of our findings, we conducted several control analyses. One of these analyses examined the win-stay, lose-shift behavior of participants, which we could confirm with the data of our experiment. The significant trends suggested that participants rather stayed with a decision when it previously led to a positive outcome. Here, it did not matter if the decision was made by the robot or themselves. Participants showed the behavioral stay-pattern in both decision agency conditions. Therefore, it seems as if they also understood the decision of Cozmo, the robot, as informative for themselves. Shifting behavior, on the other hand, was significantly influenced by decision agency. It was less likely that participants shifted their decision, when they were the decider in the first place, even when their decision led to a negative outcome. This suggests a self-bias when reconsidering a previous decision and its outcome.

Moreover, we found that participants were faster in confirming the outcome attribution towards the robot in cases where they already owned more money than the robot, suggesting a relative comfort in sharing when one already was in advantage.

Furthermore, our sample characteristics shows that the majority of individuals in our study were right-handed. Both left and right-handed participants were faster when using their right index finger, aligning with our hypothesis that keeping is easier than giving (keeping was coherently confirmed with the right index finger). Nevertheless, other factors, such as the common practice of using the right hand and index finger with a computer mouse even among left-handers, could have biased these results and the so far given interpretation. Additionally, our analysis showed that the repetition of key usage, such as using the left index finger twice to select the left item and later for giving the outcome to the robot, did not affect response times.

Our study design seemed to effectively capture very subtle effects by observing behavioral patterns in low-level action scenarios, highlighting that even simple and automated actions are influenced by psychological factors such as decision agency, outcome ownership, and outcome evaluation. However, there are also several limitations to the presented study. Recruiting specific participant groups could be insightful to further understand the phenomena presented above. Analyzing the questionnaires included in our setup, which will be part of another work, will further elucidate how personality traits are associated with cognitive dissonance in robot interactions. For instance, participants with more need for autonomy or higher depression scales might experience more discomfort when interacting with robots. Future studies should explore variations in interaction, such as engaging with another human or facing outcomes that are lost rather than given to someone else, to further validate our findings and deepen our understanding of the role of the robot's avatar in our experimental design.

In conclusion, this study provided comprehensive insights into three factors affecting cognitive dissonance in human-robot interactions. By confirming the significant roles of decision agency, outcome ownership, and their interaction with outcome value, we highlighted the complexities of designing effective human-robot interaction interfaces. Even in situations where a positive outcome generated by a robot was ascribed to a human, it seems as if the human would have experienced less cognitive conflict if he or she took the initial decision him or herself and therefore would have preferred to be the agent. Following this thought and in the light of our results, we think that human robot interactions should be designed in a way that humans always feel or experience highest possible agency. Even if a decision was automatically taken by a robot, the human should probably have the option of revise or accept it, in cases where the interaction scenarios allow for it. Our findings underline the importance of creating human autonomy and control in interactions with robots to minimize cognitive conflict and enhance the overall experience when collaborating with them.

REFERENCES

Abubshait, A., Therkelsen, S. E., McDonald, C., & Wiese, E. (2023). Forced Prosocial Behaviors towards Robots Induce Cognitive Conflict. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *67*(1), 434–439. https://doi.org/10.1177/21695067231192897

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bandura, A., Caprara, G. V., Barbaranelli, C., Gerbino, M., & Pastorelli, C. (2003). Role of Affective Self-Regulatory Efficacy in Diverse Spheres of Psychosocial Functioning. *Child Development*, *74*(3), 769–782. https://doi.org/10.1111/1467-8624.00567

Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 130–149. https://doi.org/10.1037/0278-7393.20.1.130

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Broadbent, E. (2017). Interactions With Robots: The Truths We Reveal About Ourselves. *Annual Review of Psychology*, *68*(1), 627–652. https://doi.org/10.1146/annurev-psych-010416-043958

Cassotti, M., Houdé, O., & Moutier, S. (2011). Developmental changes of win-stay and loss-shift strategies in decision making. *Child Neuropsychology*, *17*(4), 400–411. https://doi.org/10.1080/09297049.2010.547463

Chin, E., Felt, A. P., Sekar, V., & Wagner, D. (2012). Measuring user confidence in smartphone security and privacy. *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 1–16. https://doi.org/10.1145/2335356.2335358

Eaton, W. W., Smith, C., Ybarra, M., Carles, M., & Tien, A. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, 363–377.

Festinger, L. (2001). *A theory of cognitive dissonance* (Reissued by Stanford Univ. Press in 1962, renewed 1985 by author, [Nachdr.]). Stanford Univ. Press.

Fointiat, V. (2004). "I KNOW WHAT I HAVE TO DO, BUT…" WHEN HYPOCRISY LEADS TO BEHAVIORAL CHANGE. *Social Behavior and Personality: An International Journal*, *32*(8), 741–746. https://doi.org/10.2224/sbp.2004.32.8.741

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gallagher, S. (2013). Ambiguity in the Sense of Agency. In A. Clark, J. Kiverstein, & T. Vierkant (Eds.), *Decomposing the Will* (pp. 118–135). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199746996.003.0007

Gentsch, A., & Schütz-Bosbach, S. (2015). Agency and Outcome Prediction. In P. Haggard & B. Eitam (Eds.), *The Sense of Agency* (pp. 217–234). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190267278.003.0009

Giersiepen, M., Schütz-Bosbach, S., & Kaiser, J. (2023). Freedom of choice boosts midfrontal theta power during affective feedback processing of goal-directed actions. *Biological Psychology*, *183*, 108659. https://doi.org/10.1016/j.biopsycho.2023.108659

GitHub. (2008). https://github.com/

Graham, G., & Stephens, G. L. (Eds.). (1994). Mind and Mine. In *Philosophical Psychopathology* (pp. 91–109). The MIT Press.

Haggard, P., & Eitam, B. (Eds.). (2015). *The sense of agency*. Oxford University Press.

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology (2nd ed.).* (pp. 3–24). American Psychological Association. https://doi.org/10.1037/0000135-001

Hassall, C. D., Silver, A., Turk, D. J., & Krigolson, O. E. (2016). We are More Selfish than We Think: The Endowment Effect and Reward Processing within the Human Medial-Frontal Cortex. *Quarterly Journal of Experimental Psychology*, *69*(9), 1676–1686. https://doi.org/10.1080/17470218.2015.1091849

Holy, T. (2024). *Generate maximally perceptually-distinct colors* [Matlab]. https://www.mathworks.com/matlabcentral/fileexchange/29702-generate-maximally-perceptually-distinct-colors

Home—PsychoPy®. (2024). https://psychopy.org/

Jakob, H. (2007). The sense of self in phenomenology of agency and perception. *Psyche: Zeitschrift Fuer Psychoanalyse Und Ihre Anwendungen*, *13*(1), 1–20.

Jean Tsang, S. (2019). Cognitive Discrepancy, Dissonance, and Selective Exposure. *Media Psychology*, *22*(3), 394–417. https://doi.org/10.1080/15213269.2017.1282873

Kaiser, J., Buciuman, M., Gigl, S., Gentsch, A., & Schütz-Bosbach, S. (2021). The Interplay Between Affective Processing and Sense of Agency During Action Regulation: A Review. *Frontiers in Psychology*, *12*, 716220. https://doi.org/10.3389/fpsyg.2021.716220

Kotan, M. (2010). Freedom or happiness? Agency and subjective well-being in the capability approach. *The Journal of Socio-Economics*, *39*(3), 369–375. https://doi.org/10.1016/j.socec.2009.11.003

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2014). Internale-Externale-Kontrollüberzeugung-4 (IE-4). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/ZIS184

Krigolson, O. E., Hassall, C. D., Balcom, L., & Turk, D. (2013). Perceived ownership impacts reward evaluation within medial-frontal cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 262–269. https://doi.org/10.3758/s13415-012-0144-4

Leng, Y., & Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: An ERP study. *Neuropsychologia*, *48*(2), 448–455. https://doi.org/10.1016/j.neuropsychologia.2009.10.002

Levin, D., Harriott, C., Paul, N. A., Zheng, T., & Adams, J. A. (2013). Cognitive Dissonance as a Measure of Reactions to Human-Robot Interaction. *Journal of Human-Robot Interaction*, *2*(3), 1–17. https://doi.org/10.5898/JHRI.2.3.Levin

Lindstrom, M. J., & Bates, D. M. (1988). Newton—Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association*, *83*(404), 1014–1022. https://doi.org/10.1080/01621459.1988.10478693

Little, D. R., Eidels, A., Fifić, M., & Wang, T. S. L. (2018). How Do Information Processing Systems Deal with Conflicting Information? Differential Predictions for Serial, Parallel, and Coactive Models. *Computational Brain & Behavior*, *1*(1), 1–21. https://doi.org/10.1007/s42113-018-0001-9

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015). Sacrifice One For the Good of Many?: People Apply Different Moral Norms to Human and Robot Agents. *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 117–124. https://doi.org/10.1145/2696454.2696458

Mau, S., Westheuser, L., & Lux, T. (2024). *Triggerpunkte: Konsens und Konflikt in der Gegenwartsgesellschaft* (7. Auflage). Suhrkamp.

Mehta, M. M., Na, S., Gu, X., Murrough, J. W., & Morris, L. S. (2023). Reward-related self-agency is disturbed in depression and anxiety. *PLOS ONE*, *18*(3), e0282727. https://doi.org/10.1371/journal.pone.0282727

Metzger, M. J., Hartsell, E. H., & Flanagin, A. J. (2020). Cognitive Dissonance or Credibility? A Comparison of Two Theoretical Explanations for Selective Exposure to Partisan News. *Communication Research*, *47*(1), 3–28. https://doi.org/10.1177/0093650215613136

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Open Science Tools Ltd. (2019). https://opensciencetools.org/index.html#mission

Pavlovia. (2023). https://pavlovia.org/

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Platform · Sona Systems. (2002). Sona Systems. https://www.sona-systems.com/platform/

Prolific | Quickly find research participants you can trust. (2014). Prolific. https://www.prolific.com/

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459. https://doi.org/10.1111/psyp.12370

Scipy.stats.chisquare. (2024). ScyPi. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chisquare.html

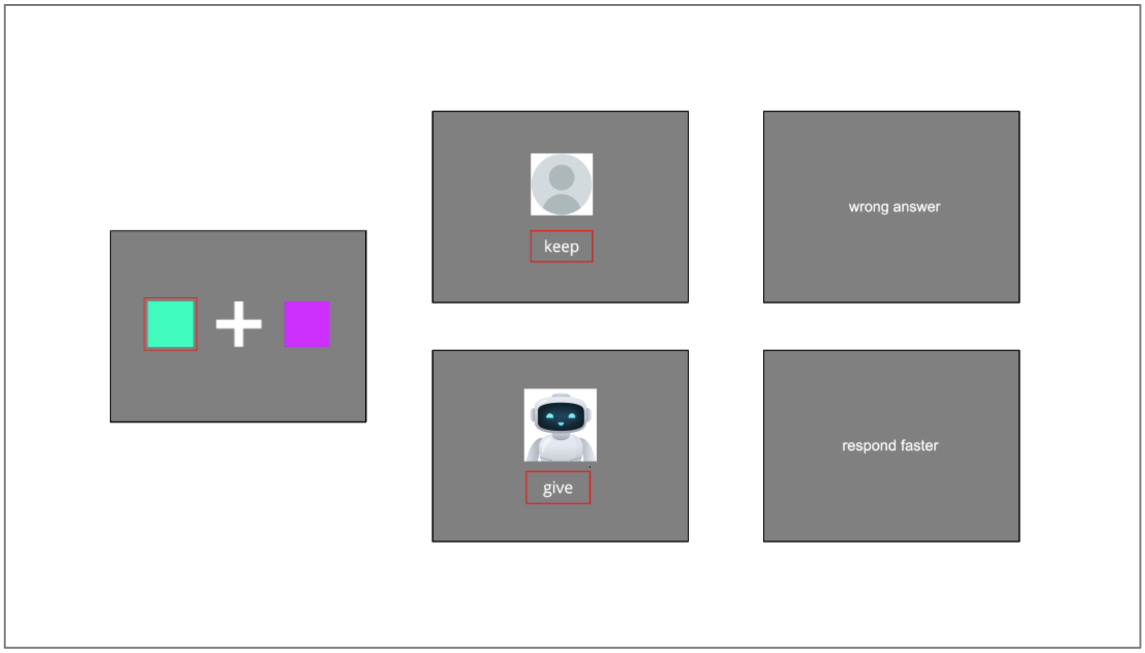
Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and Statistical Modeling with Python*. 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Stone, J., Aronson, E., Crain, A. L., Winslow, M. P., & Fried, C. B. (1994). Inducing Hypocrisy as a Means of Encouraging Young Adults to Use Condoms. *Personality and Social Psychology Bulletin*, *20*(1), 116–128. https://doi.org/10.1177/0146167294201012

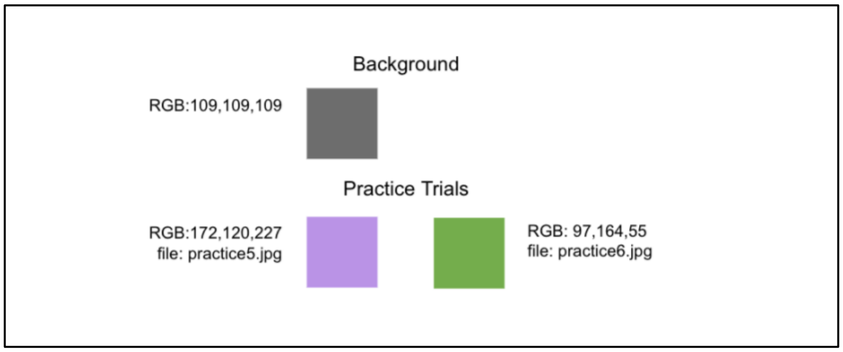
Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review*, *20*(2), 364–371. https://doi.org/10.3758/s13423-012-0324-9

APPENDIX

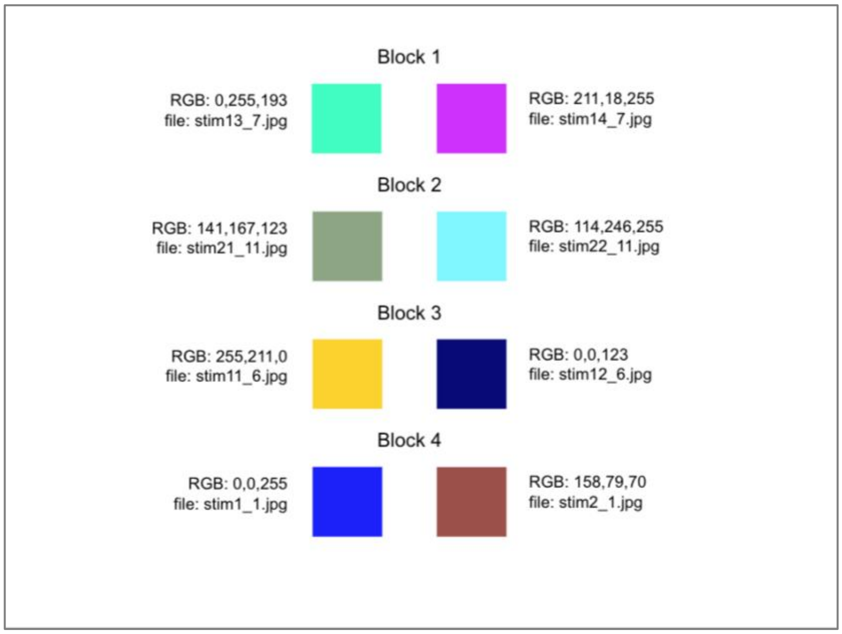
Appendix A: Breakout screens

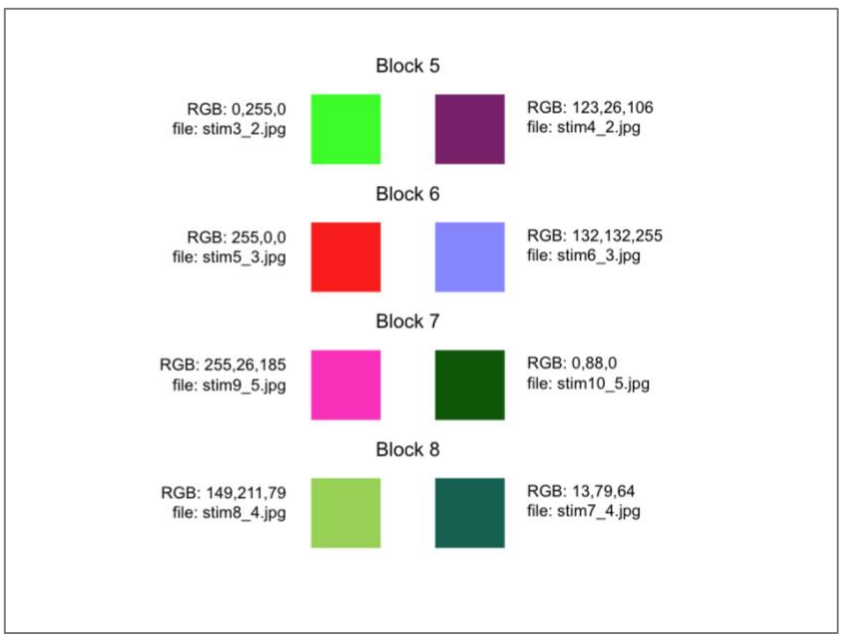


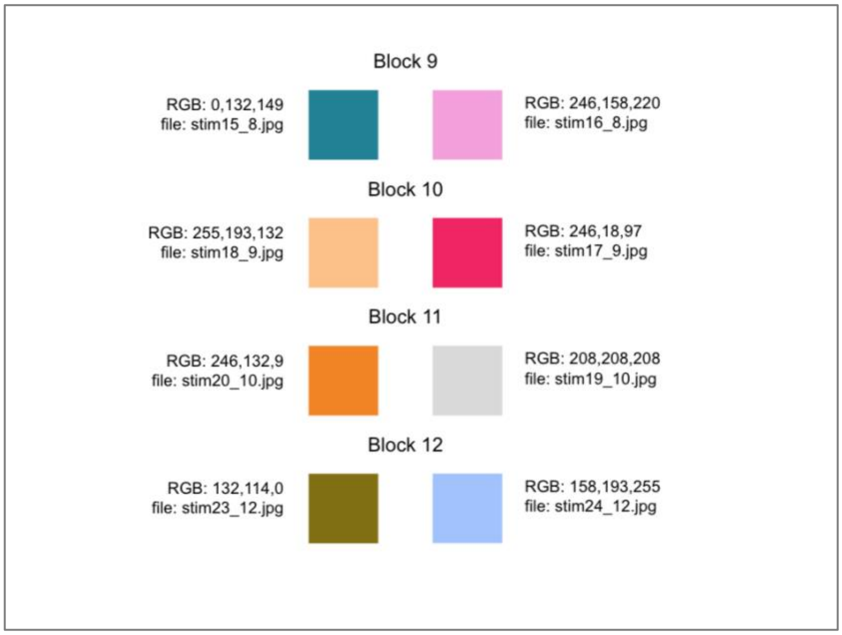
Appendix B: Colors practice trials



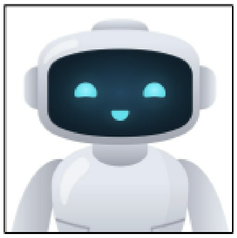
Appendix C: Colors experiment trials

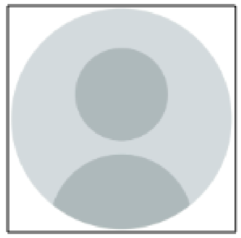





Block 9
RGB: 0,132,149
file: stim15_8.jpg
RGB: 246,158,220
file: stim16_8.jpg
Block 10
RGB: 255,193,132
file: stim18_9.jpg
RGB: 246,18,97
file: stim17_9.jpg
Block 11
RGB: 246,132,9
file: stim20_10.jpg
RGB: 208,208,208
file: stim19_10.jpg
Block 12
RGB: 132,114,0
file: stim23_12.jpg
RGB: 158,193,255
file: stim24_12.jpg

Appendix D: Avatars





Appendix E: Procedure time settings

<table>
<tr><td>5 minutes</td><td colspan="8">information about experiment, consent and survey about demographics to begin experiment; as much time as needed
approx. 5 minutes</td><td rowspan="3"></td></tr>
<tr><td rowspan="2">2,5 seconds</td><td colspan="8">info screen to display the current testing block e.g., "Block 1 of 12"; once before each block
1,5 seconds</td></tr>
<tr><td colspan="8">fixation cross; once before each trial
1 second</td></tr>
<tr><td rowspan="3"></td><td colspan="3">information "you choose" with avatar
1 second</td><td colspan="5">information "cozmo chooses" with avatar
1 second</td><td rowspan="3"></td></tr>
<tr><td colspan="3">displaying color boxes
ends with keyboard input or after max. 1 second</td><td colspan="5">white frame around one of the boxes to visualize cozmos choice
ends with keyboard input or after max. 1 second</td></tr>
<tr><td colspan="3">white frame to visualize own choice
0,5 seconds</td><td colspan="5">black or red frame to visualize choice confirmation of particpant; red in case of too slow response or wrong input, black in case of correct confirmation
0,5 seconds</td></tr>
<tr><td rowspan="4">6 seconds</td><td colspan="3">outcome value information; either 0,4€ for winning or -0,2€ for losing
1 second</td><td colspan="3">outcome value information; either 0,4€ for winning or -0,2€ for losing
1 second</td><td rowspan="4">information text "wrong answer"
1 second</td><td rowspan="4">information text "respond faster"
1 second</td><td rowspan="4">6
-
3,5
seconds</td></tr>
<tr><td colspan="3">outcome attribution information; either "give" or "keep" with the according avatar
ends with keyboard input or after max. 1 second</td><td colspan="3">outcome attribution information; either "give" or "keep" with the according avatar
ends with keyboard input or after max. 1 second</td></tr>
<tr><td colspan="3">white or red frame around "give" or "keep" to visualize outcome attribution confirmation of participant; red for too slow response or wrong input, white in case of correct confirmation 0,5 seconds</td><td colspan="3">white or red frame around "give" or "keep" to visualize outcome attribution confirmation of participant; red for too slow response or wrong input, white in case of correct confirmation
0,5 seconds</td></tr>
<tr><td>displaying both avatars and both current balances
1 seconds</td><td>information text "wrong answer"
1 seconds</td><td>information text "respond faster"
1 seconds</td><td>displaying both avatars and both current balances
1 seconds</td><td>information text "wrong answer"
1 seconds</td><td>information text "respond faster"
1 seconds</td></tr>
<tr><td>30 seconds</td><td colspan="8">continue with next block information; Countdown running; once after each block
ends with keyboard input "space" or after max. 3 minutes</td><td rowspan="3"></td></tr>
<tr><td>10 minutes</td><td colspan="8">additional Questionnaires
approx. 10 minutes</td></tr>
<tr><td>0,5 minutes</td><td colspan="8">"thanks for your participation" screen, contact details for feedback
approx. 30 seconds</td></tr>
</table>

Appendix F: Data dictionary – questionnaire data

column	explanation
age	age of the participant
gender	gender of the participant; either “female“, “male“ or “other“
handedness	handedness of the participant; either “right“, “left“ or “other“
cesdr_poor_appetite cesdr_shake_off_blues cesdr_keep_mind cesdr_felt_depressed cesdr_restless_sleep cesdr_sad cesdr_get_going cesdr_nothing_made_happy cesdr_bad_person cesdr_lost_interest_activities cesdr_more_sleep cesdr_slowly cesdr_fidgety cesdr_wish_dead cesdr_hurt_myself cesdr_tired cesdr_not_like_myself cesdr_lost_weight cesdr_trouble_sleep cesdr_not_focus	answers to questions from the CESDR questionnaire
ie4_own_boss ie4_will_succeed	answers to questions from the IE4 questionnaire

ie4_determined_by_others ie4_fate	

Appendix G: Data dictionary – response time data

column	explanation
date	date when the data collection happened
time	time when the data collection started
session	unique identifier of the session of the participant; in case one participant tries to start the experiment several times the session id raises
participant	unique identifier of the participant
block	identifier of the of block to which the trial belongs to; either “test“ or numbers from 1-12; predefined in the setting file
trial	identifier of the trial within block; numbers from 1-32; predefined in the setting file
identifier_chooser_owner_value_corr	identifier which informs about the used set of conditions within the trial; concatenated from predefined values in setting file: chooser, owner and value (e.g., cozmo_cozmo_win); corrected in preprocessing for the chance trials where the value changes to the opposite
chooser	identifies the chooser of the stimuli in the trial; either “you“ or “Cozmo“; predefined in the setting file
left_color	identifies which image from repository was used for the left square; predefined in the

	setting file; two fixed colors per block but changing their sides
right_color	identifies which image from repository was used for the right square; predefined in the setting file; two fixed colors per block but changing their sides
choice_location	gives information which of the two colored squares was chosen from participant or Cozmo; either “left“ or “right“
choice_confirm_keys	gives information which input was used to confirm Cozmos color choice, either “g“ or “k“; only filled when chooser = Cozmo; can be used to determine if participant made an error
value	identifier which informs about the outcome value of trial; predefined in the setting file; either “win“ or “lose“; corrected in preprocessing for the chance trials where the value changes to opposite
value_distribution	gives information if trial was randomly selected as chance trial or not; either 10 or 90; predefined in the setting file
owner	identifies the owner of the outcome value of the trial; either “you“ or “Cozmo“; predefined in the setting file
owner_confirm_keys	gives information which input was used to confirm owner attribution, either “g“ for give or

	“k“ for keep; can be used to determine if participant made an error
owner_confirm_rt	dependent variable; response time of participants when confirming the outcome attribution
bool_slow_color_choice_or_confirm	gives information if the participant was too slow in the decision; either when confirming Cozmo’s color choice or choosing oneself; either “true“ or “false“
bool_wrong_color_confirm	gives information if the participant was wrong when confirming Cozmo’s color choice; either “true“ or “false“; can only be true when chooser = Cozmo
bool_slow_owner_confirm	gives information if the participant was too slow in confirming the outcome attribution; either “true“ or “false“
bool_wrong_owner_confirm	gives information if the participant was wrong in confirming the outcome attribution; either “true“ or “false“
trial_index_within_block	gives information about the order of the trials within blocks

Appendix H: Mixed linear model information – two-way interactions

	model	mixed lm
	no. observations	24644
participant kept outcome	no. groups	141
	min. group size	127
effect of chooser on	max. group size	192
response time (1)	mean group size	174.80
&	dependent variable	owner_confirm_rt
effect of value on	method	REML
response time (2)	scale	0.01
	log-likelihood	17504.07 (1) & 17626.13 (2)
	converged	Yes
	model	mixed lm
	no. observations	24262
participant gave outcome	no. groups	141
	min. group size	128
effect of chooser on	max. group size	191
response time (1)	mean group size	172.10
&	dependent variable	owner_confirm_rt
effect of value on	method	REML
response time (2)	scale	0.01
	log-likelihood	17369.97 (1) & 17222.83 (2)
	converged	Yes
positive outcome	model	mixed lm
	no. observations	142398
effect of chooser on	no. groups	141
response time (1)	min. group size	139
&	max. group size	189
effect of owner on	mean group size	173.00
response time (2)	dependent variable	owner_confirm_rt

	method	REML
	scale	0.01
	log-likelihood	17043.07 (1) & 17624.35 (2)
	converged	Yes
negative outcome effect of chooser on response time (1) & effect of owner on response time (2)	model	mixed lm
	no. observations	24508
	no. groups	141
	min. group size	132
	max. group size	192
	mean group size	173.80
	dependent variable	owner_confirm_rt
	method	REML
	scale	0.01
	log-likelihood	17129.70 (1) & 17012.86 (2)
	converged	Yes
participant chose effect of owner on response time (1) & effect of value on response time (2)	model	mixed lm
	no. observations	24610
	no. groups	141
	min. group size	139
	max. group size	192
	mean group size	174.50
	dependent variable	owner_confirm_rt
	method	REML
	scale	0.01
	log-likelihood	17673.57 (1) & 17448.85 (2)
	converged	Yes
Cozmo chose	model	mixed lm
	no. observations	24296
	no. groups	141
	min. group size	131

effect of owner on response time (1) & effect of value on response time (2)	max. group size	191
	mean group size	172.30
	dependent variable	owner_confirm_rt
	method	REML
	scale	0.01
	log-likelihood	17069.88 (1) & 16813.87 (2)
	converged	Yes

Appendix I: Mixed linear model regression results of value and chooser on stay ratio

model:	mixedlm	dependent variable:	ratio
no. observations:	564	method:	REML
no. groups:	141	scale:	0.00
min group size:	4	log-likelihood:	1006.50
max group size:	4	converged:	yes
mean group size:	4		

Appendix J: Mixed linear model regression results of value and chooser on shift ratio

model:	mixedlm	dependent variable:	ratio
no. observations:	560	method:	REML
no. groups:	141	scale:	0.00
min group size:	3	log-likelihood:	1015.63
max group size:	4	converged:	yes
mean group size:	4		

Appendix K: Mixed linear model regression results of balance on response times

model:	mixedlm	dependent variable:	owner_confirm_rt
no. observations:	24262	method:	REML
no. groups:	141	scale:	0.01
min group size:	128	log-likelihood:	17218.31
max group size:	191	converged:	yes
mean group size:	172.10		

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.58	0.01	104.02	0.00	0.57	0.59
participant_owns_more_than_cozmo[true]	-0.01	0.00	-4.87	0.00	-0.01	-0.00
group Var	0.00	0.00				

Appendix L: Mixed linear model regression results of handedness on response time

model:	mixedlm	dependent variable:	owner_confirm_rt
no. observations:	48540	method:	REML
no. groups:	140	scale:	0.0141
min group size:	275	log-likelihood:	34152.66
max group size:	381	converged:	yes
mean group size:	346.7		

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.59	0.02	33.04	0.00	0.56	0.63
handedness [right]	-0.02	0.02	-0.10	0.32	-0.06	0.02
owner_confirm_keys [k]	-0.04	0.00	-10.72	0.00	-0.05	-0.03
handedness [right]: owner_confirm_keys [k]	0.00	0.00	0.94	0.35	-0.00	0.01
group Var	0.00	0.00				

Appendix M: Mixed linear model regression results of repetition on response time

model:	mixedlm	dependent variable:	owner_confirm_rt
no. observations:	48906	method:	REML
no. groups:	141	scale:	0.0144
min group size:	275	log-likelihood:	33984.97
max group size:	381	converged:	yes
mean group size:	346.90		

	coef.	*std. err.*	*z*	*p>|z|*	*[0.025*	*0.0975]*
intercept	0.56	0.01	105.95	0.00	0.55	0.57
location_decision_owner_same[true]	0.00	0.00	0.21	0.83	-0.00	0.00
group Var	0.00	0.00				