MASTER

Delegating Subjective Decisions to AI
The Effect of Choice Characteristics

van den Bergen, Doortje

Award date:
2023

21-4-2023

Delegating Subjective Decisions to AI: The Effect of Choice Characteristics

By Doortje van den Bergen

0993592

in partial fulfillment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:
dr. C. Zhang
dr. M.C. Willemsen
(dr. D. Lakens)

Acknowledgments

This work would not have been possible without my two supervisors Dr. Chao Zhang and Dr. Martijn Willemsen. I would like to thank them for their excellent feedback, understanding, and inspiring words. In addition, I would like to thank my parents for always being there for me and for always believing in me. Most importantly, I would like to thank my boyfriend David for his mental support and for helping me work through my most challenging programming issues. I am also grateful for all my friends and family who were willing to participate in my everlasting pre-tests.

Abstract

The delegation of subjective choices to AI is becoming more common. Whereas prior research on decision-making revealed an effect of choice difficulty on delegation, the effect of choice characteristics is largely understudied. The present study aims to fill this gap and compares people's delegation rates to AI across five decision situations including two-option two-attribute choice sets. We compare situations in which: (1) one option dominates the other; (2) two options are equally bad; (3) two options are equally good. Furthermore, we look at trade-offs in which; (1) one option has attribute values that are close together; (2) the attribute values of the two options are far apart. We conducted an online study (N=96) measuring delegation rates in two contexts including the choice between two hotels and two monetary gambles. Moreover, we gathered qualitative data on people's reasons for delegating to AI, as well as on their adopted choice strategies. The results revealed that for the hotel context, people delegate significantly less in the dominant vs. the other decision conditions. For the gambling context, people delegated significantly less in the dominant and first tradeoff condition vs. the other conditions. Additionally, delegation rates in the second tradeoff condition were significantly higher than in the equally good and bad conditions. Overall, people delegated significantly less in the hotel vs. the gambling context. Our qualitative data demonstrate that our findings are largely attributable to choice difficulty, in addition to trust and personal differences (e.g. regarding risk aversion, anticipated regret). We discuss the relevance of these findings to the decision-making literature, as well as implications for future research and practical ones for designers of AI-driven delegation systems.

Contents

1. Introduction

In our everyday lives, we are confronted with many difficult decisions. Particularly for subjective choices, we may experience indecisiveness. Indecisiveness in decision-making can have different implications, such as delaying one's choice, for example, to gather more information or to look for alternatives, not choosing at all, tunnel vision (e.g. choosing the default option), and regret (Rassin, 2007; Shafir et al., 1993). It might occur because of a variety of reasons, such as the cognitive or emotional complexity of a decision (e.g. Larrick, 1993; Luce et al., 1997), uncertainty of how our actions might turn out (Shafir & Tversky, 1992), or anticipated regret (Steffel & Williams, 2017).

From the decision-making literature, we already know that people may forego the negative consequences of indecisiveness by delegating their decisions to other people. In general, people are more likely to avoid and delegate difficult decisions compared to easy ones (e.g. Anderson, 2003; Steffel & Williams, 2017). Aside from difficulty, some factors playing a role in delegation to other people are people's perceived stakes, perceived trust, and one's own level of expertise (e.g. Gur & Bjørnskov, 2017; Leana, 1986; Logg et al., 2019).

We may not just delegate difficult decisions to others but may also opt for a full-choice delegation to AI. This is also becoming more common given the rapid development of AI and popular usage of apps like ChatGPT (Brockman et al., 2023). A full delegation implies that AI does not just inform our decisions, but that it acts as a decision-maker on our behalf. Like recommender systems, it may make predictions based on our personal preferences and prior decision-making processes. It is however different from recommender systems, as these do not act as decisionmakers but only as recommenders. The process of full delegation to AI is also known as automatic decision making (ADM), which includes a completely automated decision-making process in which there is no human involvement (Harris & Davenport, 2005).

Earlier studies on the delegation to AI have shown that similar factors may play a role as in delegation to people (e.g. perceived trust, levels of expertise). However, little is known about the effect of choice characteristics on the delegation to AI. As mentioned before, delegation to others largely depends on choice difficulty. We already know that this is largely affected by the characteristics of a choice set. Shafir et al. (1993) indicate that decisions including tradeoffs may especially be difficult compared to decisions with a dominant alternative. This seems to be due to increased cognitive difficulty (Bettman et al., 1993). Additionally, research suggests that decisions between options that we deem equally attractive and similar may be easier than decisions including tradeoffs (Kim et al., 2013). In this case, these similar options are likely to be seen as substitutes for reaching the same goal, meaning that participants may be indifferent between two options (e.g. Xu et al., 2013).

To expand on this topic, the present study examines the effect of choice characteristics on the delegation of subjective choices to AI, while combining the findings on trust in AI with those from the fields of choice, delegation, and the control premium. Specifically, we will focus on whether people's delegation rates to AI are different for five qualitatively different decision situations. Firstly, we examine situations in which: (1) one option dominates the other; (2) two options are equally bad; (3) two options are equally good. In addition, we examine trade-offs in which; (1) one option has attribute values that are close together; (2) the attribute values of the two options are far apar. The situation with a dominant alternative usually implies an easy decision. The other four are all difficult but with different types of conflicts.

We measure this difference for two decision contexts; one that entails personal, hypothetical scenarios regarding the choice of a hotel, and one that entails gambling decisions involving real monetary consequences. We present our hypotheses in Section 3 and support them with existing findings from the fields mentioned above. Subsequently, we provide an overview of the method and stimuli design of the experiment designed to test our hypotheses for the two decision contexts, followed by its results. In conclusion, we discuss our results by describing theoretical implications and practical ones for both users and companies that use ADM for personalized offerings.

2. Related Work

2.1. The effect of choice characteristics on choice difficulty and preferences

2.1.1. Tradeoffs in choice

A tradeoff implies that a gain in one choice attribute is offset by a loss on the other. First, imagine making a tradeoff between when choosing between two hotels, that differ with regards to the time travelling to the city center and their price per night. An example of a small tradeoff is when a difference in price of €5,00 is offset by a 7 min. difference in travelling time. We can also quantify the size of this tradeoff by defining the *exchange rate* between attributes. That is, the tradeoff size in the small tradeoff condition can be quantified as $(€5/7) *$ time = € 0,71/time, meaning that an increase of 71 cents is compensated by one minute less travelling time. For other tradeoffs, the exchange rate can be equal, although the tradeoff itself might be larger. For instance, the exchange rate in a large tradeoff condition can be quantified as $(€15/21) *$ time = € 0,71/time. However, the difference between the attributes is larger, determining tradeoff size.

Again, research has shown that especially decisions requiring large trade-offs are considered more difficult, both (emotionally and cognitively) than decisions including moderate to small trade-offs (Chatterjee & Heath, 1996; Bettman et al., 1993). In line with this, Scholten and Sherman (2006) have proposed a double-mediation model, describing the relationship between tradeoff size and choice difficulty. This model has an inverted-U-shape, and describes that for very small tradeoffs, people experience less choice difficulty. Again, they might experience maximal choice difficult for moderate tradeoffs, and less choice difficulty for large tradeoffs. They assumed that the effect of tradeoff size on choice difficulty in the upward direction is due to the greater sacrifices that are inherent to choosing one option instead of another. The downward-shaped effect was ascribed to the mediating effect of argumentation. Put differently, people seem to find it easier to justify their choices to themselves when the tradeoff size is large.

Other research has also shown an increased choice difficulty for large trade-offs. Chatterjee and Heath (1996) for instance, studied choices between two cameras or two job candidates, while employing a 2x2 design (small vs. large tradeoff, cameras vs. job candidates). They found that participants perceived choices in the large tradeoff condition as more difficult. In addition, Dhar and Simonson (2003) have reported that people are more likely to defer difficult trade-offs. In their research, they focused on hypothetical purchase decisions in different product categories. On top of that, they examined the effect of a no-choice option on the compromise effect. In short, this effect arises when a third option is added to a two-option two-attribute choice set. This set may initially include option a, with extreme values on the first attribute, and option b, with average values for both attributes. After adding a third-choice option with extreme values for the second attribute, the choice chare of option b, relative to a, is expected to increase.

In pursuit of studying this effect, the authors presented participants with a two-option two-attribute choice. Half of the participants also had a no-choice option. They found that when a no-choice option was available, an option that was average on all attributes (the compromise option) was chosen less often than an option that was the best on one of the attributes. The authors have attributed this difference to a decrease in the compromise effect: people who find it hard to make tradeoffs usually go for the compromise option but find choice deferral more attractive when it is available.

2.1.2. Perceived attractiveness vs. perceived similarity

In general, choice options can be perceived as equally good, which is often referred to as an *approach-approach* conflict (e.g. Arkoff, 1957; Chatterjee & Heath, 1996). They can however also be perceived as equally bad, commonly referred to as an *avoidance-avoidance* conflict (e.g. Arkoff, 1957; Chatterjee & Heath, 1996). Following *reference dependence theory* (Arkoff, 1957; Tversky & Kahneman, 1991), attribute values that are more positive in value result in a positive perception or perceived gains. In the equally good choice set in Table 1 for instance, all cleanness and quietness values are above the natural 3.5 reference point (0= lowest, 7=best). The exception is the equally bad choice set, in which all values are negative in comparison to the reference, thus implying losses. That is, all attribute values for cleanness and quietness are below 3.5. Nonetheless, choice sets may also imply an *approach-avoidance* (or mixed valence) conflict, as is the case when one of the attributes has negatively perceived values, and the other has positively perceived values.

Table 1.

Example of the equally good and bad stimuli for the hotel context

	Cleanness	**Quietness**
	Equally good	
Hotel A	4.2	7.0
Hotel B	4.3	6.9
	Equally bad	
Hotel A	0.8	2.0
Hotel B	0.9	1.9

Different from the perceived similarity in attractiveness, choice options may also have an overall *perceived similarity.* It seems that this is not just based on perceived equal attractiveness but may also be based on two different mental presentations. First, people may have a concrete representation of their choices, and base the perceived similarity between options on the difference between the different attribute values (Xu et al., 2013). Second, Xu et al. (2013) have shown that some people may base perceived similarity on commonalities between options and on how well options serve goal fulfillment. More specifically, they found that people with more abstract representations (based on goal fulfillment) perceive the choice options in a large assortment as more similar than those with a concrete representation of the assortment.

Independent of mental representations, the literature has found mixed results regarding whether the increased similarity between options decreases choice difficulty (Kim et al., 2013; Willemsen et al., 2016). Kim et al. (2013) showed that the perceived similarity of two choice options decreased choice difficulty. Plus, it had a negative effect on choice avoidance. They showed that for two products, one can increase perceived similarity when introducing small differences in attributes that would otherwise be similar. In their study, participants were allocated to a same-price or different-price condition. In the same-price condition, they had to choose between two teas (lemon ginger or citrus cinnamon) with the same price. In the different-price condition, they had to choose between one type of tea priced at \$3.68 and another one priced at \$3.78 (and vice versa). Participants in the different-price condition perceived the two products as more similar than in the same-price condition. They also found these results for other products (e.g. cereals).

Willemsen et al. (2016) reported an opposite effect and showed that only a small diverse list (implying a tradeoff between attributes) with five movie recommendations was perceived as leading to less choice difficulty and enhanced choice satisfaction compared to a list with more similar options. We will elaborate on these mixed results in the next section, which is relevant when formulating our hypothesis regarding choice delegation for choice sets with equally bad and good options.

Finally, the literature suggest that people find choices between equally good options more difficult than choices between options with a dominating alternative. In particular, Dhar (1977) have shown that the preference for a no-choice option increases after adding an equally attractive product to a choice set with one product. They tested this preference in an experiment with four conditions representing the choices regarding: 1) one alternative 2) two equally attractive alternatives differing on two attributes 3) two equally attractive alternatives differing on 4 attributes 4) one inferior and dominant alternative. As a side note, condition 2 and 3 did not differ with respect to a preference for the no-choice option. Dhar (1977) also found support for the hypothesis that decision processes resulting in this no-choice options resulted in a greater number of total thoughts, which relates to a greater choice difficulty. In addition, they found that for these processes, people have a relatively equal numbers of favorable evaluations toward each option. Lastly, they found evidence for a mediating effect of preference uncertainty on choice deferral. Phrased differently, equally attractive alternatives led to indifference about which outcome is obtained, which increased choice deferral.

2.1.3. The prominence effect

Aside from perceived attractiveness and similarity, there exists a general *prominence effect*, which may decrease choice difficulty (Fischer et al., 1999). This effect implies that people will most often choose the option that has the highest value on the attribute that is most important to them. In healthcare, for instance, it has been shown that people may choose treatment options that had a lower health risk, even though they have previously expressed indifference between those options and others that were better on the cost dimension (Persson et al., 2022). Specifically, mixed-valence sets may also spur the *negative-based prominence effect* (Willemsen & Keren, 2002). In other words, research has shown that people may most often choose the option that is least negative on the attribute with a negative valence, implying a dominant alternative (Willemsen & Keren, 2002). Again, it is important to consider that both effects may play a role in whether individuals may perceive an option as more dominant, which may decrease choice difficulty. This may in turn lead to decreased delegation rates.

2.2. Delegation to AI as a strategy to overcome indecisiveness

The previous section argues for the effect of choice characteristics on indecisiveness. Alas, there is, to our knowledge, no research that directly shows the effect of indecisiveness on delegation. Yet, findings from research on recommender systems and choice difficulty seem to suggest that this effect is likely to exist. This was also argued by Broniarczyk and Griffin (2014) in a review of decision difficulty in the era of consumer empowerment. They propose that delegation might alleviate the indecisiveness consumers may face when experiencing choice difficulty due to an ever-increasing number of purchase alternatives.

Rassin (2007) attempted to find a context-free definition of indecisiveness and defines it as "the experience of decision problems (i.e., lack of information, valuation difficulty, and outcome uncertainty) resulting in overt choice-related behaviors such as delay, tunnel vision, and post-decision dysfunctional behavior (p.10)." Indeed, existing work shows evidence for the effect of valuation difficulty on indecisiveness and choice delay. This is also highlighted by the *choice-overload* hypothesis coined in the literature (e.g. Iyengar & Lepper, 2000; Scheibehenne et al., 2010). Its basic proposition is that having more options to choose from may eventually make us less happy than having a small, limited number of options. This is because of several reasons, such as cognitive effort due to the number of item pairs to compare, and effort resulting from a choice from assortments with high density, which is the average distance between two adjacent attribute levels (Fasolo et al., 2009). Similarly, following a comprehensive literature review, Anderson (2003) concludes that selection difficulty plays a role in choice delay.

Findings by Steffel and Williams (2017) provide more empirical evidence for why people might delegate. One of their experiments looked at four conditions; people either saw a large or small selection of teas and were either in the presence of a salesperson or not. They were then asked to either purchase, opt out, or delegate their choice of tea to a salesperson if she was present. It appeared that only for large selections, people were more likely to purchase tea if they could delegate their choice. This supports the idea that people may delegate to avoid indecisiveness.

It is likely that people also find delegation to AI an appealing strategy to overcome indecisiveness. Yet, to our knowledge, there is no direct research supporting this. Still, research on recommender systems suggests that they can decrease choice difficulty if they effectively increase the diversity of a choice set (Long et al., 2022; Willemsen et al., 2016). As in ADM, these systems may predict users' preferences based on behavioral data.

As mentioned before, Willemsen et al. (2016) showed that people experienced less choice difficulty and experienced enhanced satisfaction when a movie recommender system presents small sets with highly diversified options compared to large sets with more similar options. They mainly ascribed this effect to the density principle discussed by Fasolo et al. (2009).

Namely, small diverse choice sets seem to have a smaller perceived density, reducing cognitive difficulty as the items appear less similar. If recommender systems may reduce users' cognitive effort by providing them with a small and attractive diverse set of items, it is not farfetched that if users have no strong choice preferences, they may order AI to choose one (un)attractive alternative on their behalf. This may then reduce their cognitive effort and enhance their choice satisfaction, just as choosing from a smaller list is less effortful than choosing from a large list.

2.3. Other factors playing a role in choice delegation: perceived trust, control, and stakes

2.3.1. Trust in AI

Apart from choice characteristics, a growing body of research has focused on the role of trust in the delegation to AI. Earlier work on trust in AI mostly focused on the role of advisory AI, in contrast to ADM which was then still in its infancy. The first studies revealed a general "algorithm aversion", meaning that people are generally averse to taking advice from algorithms (Dawes et al., 1989). More recent research supports this notion (Dietvorst et al., 2015). Yet, this finding appears to be more nuanced. Generally, people indeed seem to be averse to taking advice from both AI and humans, preferring to rely on their own judgments. Recent studies on algorithm appreciation, however, showed that for forecasting tasks, people may tend to rely more on advice from AI than from a person (Logg et al., 2019).

Yet, this appreciation has not been shown for the full delegation of subjective tasks to AI. For these tasks, it still holds that people prefer to rely on their own judgments (Castelo et al., 2019). In addition, they seem to rely less on delegation to algorithms than to other people (Logg, 2017). Despite of this, this aversion seems to be variable and may depend on people's perceptions of AI. Castelo et al. (2019) for instance found that increasing an AI's perceived affective human likeness diminished the effect of task objectivity in people's reliance on AI. Furthermore, aversion may diminish if people believe the AI takes their opinions into account, or is perceived to share a similar personality or thinking process (Al-Natour et al., 2011; Kawaguchi, 2021). Lastly, aversion also depends on personal factors; it might be higher for people that are older, who are less familiar with technology, or for who are extraverted as opposed to introverted (Mahmud et al., 2022). In conclusion, it appears that although people are generally averse to fully delegate subjective tasks, this aversion may be context and person dependent. This might explain why in reality people still delegate subjective choices, like the purchase of their groceries or choice of a movie, to AI.

2.3.2. The control premium

As discussed before, research on the effect of choice characteristics on delegation to AI is quite limited. We do know of one study conducted by Candrian and Scherer (2022), that has focused on that the effect of decision outcomes, namely losses and gains, on people's willingness to delegate an estimation task to AI or other people. This study revealed that for AI, people were equally willing to delegate their choices of decisions involving losses or gains. This was also found in earlier research by Bobadilla-Suarez et al. (2017). For delegation to other people, however, they found that people were more likely to delegate decisions involving gains than losses. The authors explained this through the *control premium*. This refers to the fact that participants often chose to pay more to make their own choices than they should if their aim is to maximize their pay-off. It appears that this is not driven just by overconfidence or a lack of information, but by a need for control (Bobadilla-Suarez et al., 2017).

Owens et al. (2014) also found evidence for this control premium and found it might be independent of people's aversion of uncertainty. Furthermore, this premium seems to exist because of loss aversion and fear that other agents act out of self-interest (related to trust), which poses a risk to the decisionmaker (Bohnet & Zeckhauser, 2004; Butler & Miller, 2018). As argued by Candrian and Scherer (2022), people experience AI as less self-interested, and hence, they may equally delegate decisions involving losses and gains to AI. In the next section, we turn to this theory to hypothesize whether people will have higher delegation rates for decisions between options that are equally bad or good.

2.3.3. Stakes

Lastly, it is likely that people's reliance on the delegation to AI may depend on what is at stake. Research on trust in AI for example found that people consider AI to be less trustworthy and competent when the stakes are high, such as in the case of a prison sentence, than when they are low as in the case of meal planning (Ashoori & Weitz, 2019). Classic findings from the delegation literature also found this effect for delegation to people, disclosing that managers are less likely to hand over decisions when they are more important (Leana, 1986; Yukl & Fu, 1999).

As discussed, the effect of choice characteristics on the delegation to AI is scarce, and the present research aims to fill this gap. In the next section, we hypothesize whether there is a difference in delegation rates to AI for choices with different choice characteristics, drawing on the findings discussed above. Furthermore, we look at whether these differences may differ for the two decision contexts we examine; one that entails personal, hypothetical scenarios regarding the choice of a hotel, and one that entails gambling decisions involving real monetary consequences. Before we do this, however, we review the literature on preference elicitation in Section 2.4, and more specifically that on matching, attribute weights and utility theory. This is relevant as it provides the necessary theoretical background to the method of our study.

2.4. Preference elicitation

It is intrinsic to choice that not everyone may find two options equally attractive. Amongst others, this might be due to individual differences in terms of attribute importance. When choosing a hotel for instance, one might place more importance on cleanness instead of quietness. Furthermore, people might have different exchange rates for which they are indifferent between two options[1]. We define indifference as being unable to decide between options, because they are both deemed equally attractive, or equally similar. To control for these individual differences, one could measure indicators of personal preferences through a point-allocation and matching procedure.

[1] In practice, indifference may have multiple definitions. For example, indifference may relate to a disinterest in making a choice. Yet, it may also relate to the inability to make a choice, which might stem from the similarity between options. For a further discussion on this topic, see the work by Willemsen and Keren (2003).

2.4.1. Point-allocation

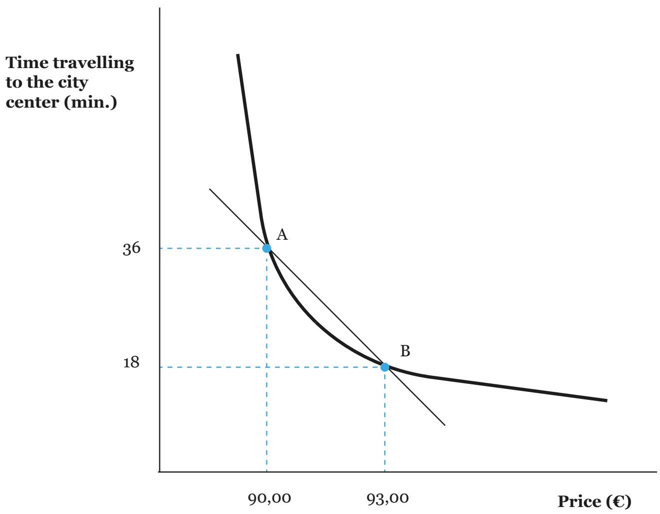
Researchers commonly apply different methods for determining indifference, or dominance of one of the options. Some research has focused on measuring attribute importance. In behavioral sciences, the point-allocation task is commonly used to measure attribute importance. This task entails dividing 100 points among different choice attributes, in which a higher number of points indicates a higher importance. Van Ittersum et al. (2007) have demonstrated its validity for measuring attribute *relevance*, proposed by Myers and Alpert (1977) to measure importance based on personal values and desires. Van Ittersum et al. (2007) however found that attribute relevance does not always reflect the importance of attributes in judgment and choice, referred to as attribute *determinance* (Myers & Alpert, 1977). Yet, the point allocation task tends to measure determinance if it is used directly after presenting information about the attribute-ranges (Van Ittersum et al., 2007).

2.4.2. A matching approach

Another common method for inferring preferences is the construction of an indifference curve. This curve is created by plotting the different exchange rates for which one is indifferent between two options. As an example, Figure 1 displays a potential indifference curve for the hotel context in our study, which in short, is about a tradeoff between two different hotel attributes. From this graph, one can infer an indifference ratio. This ratio is estimated by taking the slope of the graph in a “region” of interest. The indifference ratio would be €$(36 - 18)/(93{,}00 - 90{,}00)min.$ = €6/min., meaning that one is indifferent if a tradeoff involves trading a price of 6 euros against one minute of time travelling to the city center.

Figure 1.

Example of an indifference curve for a tradeoff between time and price in the hotel context



The matching task can be used to determine the exchange rates represented in an indifference curve. In contrast to the point-allocation task, this task is proposed to directly measure attribute *determinance* and may thus directly inform one about people's actual choice behavior. Apart from matching, marketing studies often use other methods such as the conjoint method (Louviere & Islam, 2008; Van Ittersum et al., 2007). However, this may not always be practical as it requires fitting multiple linear regression models to an individual's rating of choice attributes, requiring many data points (Louviere & Islam, 2008). Instead, the matching procedure requires participants to simply fill out a value for one attribute of an option, for which they find themselves indifferent between the options.

Further research on matching has shown that in general, people are reluctant to match on values that have a negative valence in comparison to the values of the attribute for the other option (Willemsen & Keren, 2002). Hence, it can be beneficial to employ a matching procedure for which people only match in the positive direction, meaning that they will always fill out a number for an attribute that has a more positive valence than the number it has for the other option. Table 2 provides an example of this procedure. As one can see, the matching task is based on a reference choice set, which concerns a choice between hotels with different prices and times that are required to travel to the city center. If we assume that a longer travelling time and a higher price are perceived as negative, upward matching requires matching on X1 and Y1.

Table 2.

Example of an upward matching procedure

	Attribute A	**Attribute B**
	Price	Time traveling to the city center
Hotel A	X1	45 min.
Hotel B	€ 93.00	Y1
	Reference choice set	
Hotel A	€ 90.00	45 min.
Hotel B	€ 93.00	18 min.

2.4.3. Utility theory

Finally, one could draw on *expected utility theory* (EUT) to predict for what attribute values people would be indifferent between two options in the equally good, bad and tradeoff conditions. EUT is commonly used in economics, and postulates that people evaluate choices based on the likelihood of certain outcomes. We will now touch upon some of its essential principles. First, according to EUT, people are expected to choose the option with the highest expected utility. This expected utility of a lottery $L = (p_1, \dots, p_n)$ is commonly defined by Equation 1.

$$E(u|L) = \sum_{j=1}^{N} p_j \, u(x_j) \tag{1}$$

Here, the function $f = u(x)$ assigns a certain utility to each certain outcomes $X = (x_1, \ldots, x_n)$ which may occur with a certain probability specified in the lottery L. EUT then assumes that people will only prefer lottery L over lottery L' if:

$$\sum_{j=1}^{N} p_j\, u(x_j) \geq \sum_{j=1}^{N} p'_j\, u(x_j) \qquad (2)$$

As an example, assuming $u(x) = x$, EUT expects that people prefer a lottery $L = (0.5, 0, 0.5)$ with outcomes $X = (5, 10, 20)$ over a lottery of $L' = (0, 1, 0)$ with the same outcomes $X = (6, 10, 14)$. For lottery L, $E(u|L) = 0.5(5) + 0.5(20) = 12.5$, which is bigger than $E(u|L') = 10$.

The standardly used utility functions, however, are either $u(x) = \sqrt{x}$ or $u(x) = \log(x)$ (Damodaran, 2007; Petrović et al., 2003). Their standard properties are *decreasing absolute risk aversion*- individuals will invest larger dollar amounts in risky assets when they get wealthier – and *constant relative risk aversion* – individuals will invest the same percentage of wealth in risky assets as they get wealthier (Damodaran, 2007). The square-root also seems to be appropriate for calculating utilities of subjective properties (Galanter, 1962). We have therefore used this function to compare the expected utilities of the different options in the choice tasks of this study. EUT usually only tells us something about the preference over options that have outcomes on only one attribute such as money or time. However, we also study choices between hotels with two attributes. Therefore, we also consider multi-attribute utility theory (MAUT). This theory is an expansion of EUT and proposes that a utility function $u(x_1, \ldots x_n)$ can be reduced to the additive form (Dyer, 2016):

$$\sum_{i=1}^{N} k_i\, u_i(x_i) \qquad (3)$$

where $\sum_{i=1}^{N} k_i = 1$, in which k_i represents the importance of the different utility functions for each attribute.[2] In Section 4.3 we will further elaborate how we constructed the different choice sets for each context based on this equation and the theory discussed above.

3. Research question and hypotheses

To recap, this study looks at the differences in delegation to AI for five qualitatively different decision situations, which are visualized in Figure 2 and listed in Table 3. This table also shows the assumptions that are made for each condition. In short, our research question is:

RQ. How do people's average delegation rates to AI differ across five different decision conditions with distinct choice characteristics?

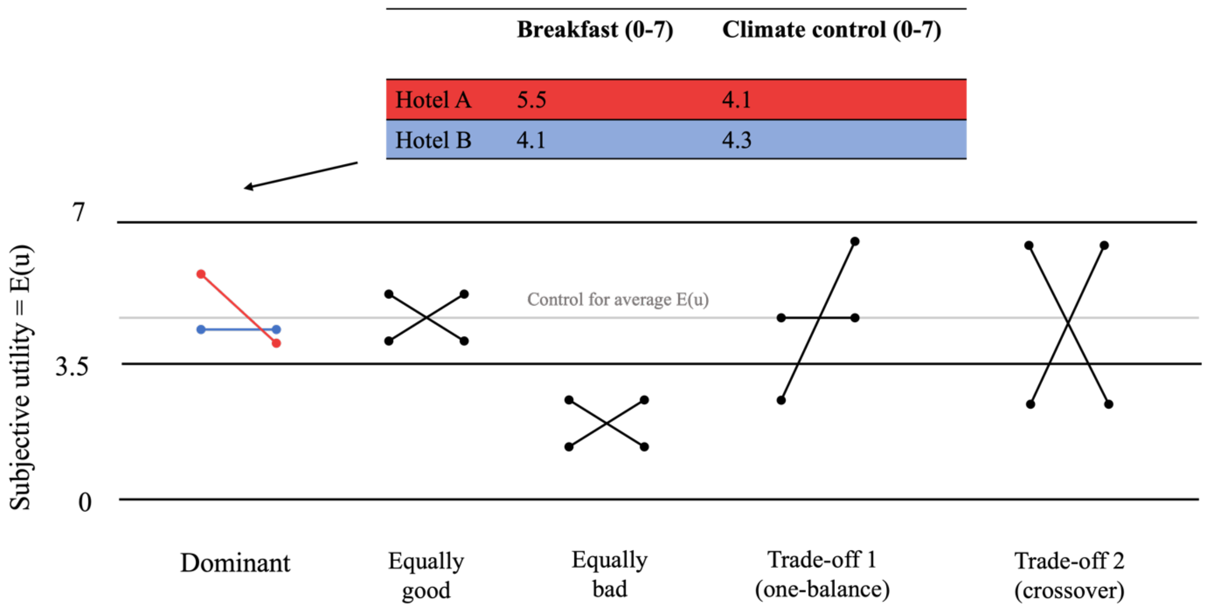
To answer our main research question, differences in delegation rates between the conditions are measured for both the coin and hotel context. For both contexts, the choice sets have similar characteristics as those described in Table 3. Therefore, the hypotheses we formulate in this section counts for both the hotel and monetary gamble context.

[2] This only holds if the marginality condition is met; this means that the preference for any lottery should only depend on the marginal probabilities of the values of an attribute, and not on their joint probabilities (Dyer, 2016).

As mentioned before, all conditions in the present research include two-option two-attribute choices. One limitation of this approach is that this represents a simplified representation of choices in the real world, in which we might have to choose between options with more than two attributes. Yet, results from a first attempt at studying delegation rates for options with two attributes may provide us with a theoretical foundation for more elaborate work. Furthermore, all conditions imply an approach-approach or an avoidance-avoidance conflict, meaning that one must choose between two options with attribute values that are all either above or all below a natural reference point. The dominant condition is an exception to this as dominance implies that there is no decision conflict; one option is clearly more attractive than the other. We have purposely excluded choice sets with mixed valence as these sets may spur the *negative-based prominence effect* (Willemsen & Keren, 2002). As our aim is to also study choice sets for which two alternatives are equally (un)attractive, this effect could confound our results for these conditions.

Figure 2.

Abstract visualization of the attribute values and the E(u) of the different choice conditions in the hotel context



Note. Each line represents one choice option, and each dot represents the option's value on one of its characteristics. The general pattern in this figure also counts for the coin context, although its expected utilities for the conditions are of a different order.

In contrast to Chatterjee and Heath (1996) we also focus on two types of relatively large tradeoffs instead of one for both decision contexts we study. Specifically, we first focus on tradeoffs that include one option that has attribute values that are close together and another option that has values that are further apart. Furthermore, we focus on tradeoffs for with both options have values that are far apart. The tradeoff size is equal for both tradeoffs. Lastly, Table 1 already provided an example of our presented choice sets in equally bad and good conditions for the hotel context. Notably, the attribute differences for these types of decisions are relatively small, implying a very small tradeoff.

Table 3.

Assumptions for the five different decision conditions

	Condition	Description	Options are equally (un)attractive	Tradeoff size
1	Dominant	Both options have attribute values that are above a natural reference point.	No	Not applicable (there is no tradeoff)
2	Equally bad	Both options have attribute values that are below a natural reference point.	Yes	Small
3	Equally good	Both options have attribute values that are above a natural reference point.	Yes	Small
4	Tradeoff 1	One option has attribute values that are close together and another option has values that are further apart. All attribute values are above a natural reference point.	Yes	Relatively large
5	Tradeoff 2	Both options have attribute values that are far apart. All attribute values are above a natural reference point.	Yes	Relatively large

3.1. Hypotheses for tradeoff conditions 1 and 2

As mentioned before, research has shown that we might have difficulty trading-off two attributes against each other (Shafir et al., 1993). Existing research suggests that this is because making a trade-off is cognitively complex, as it requires weighed addition and careful consideration of all information. In contrast, simple heuristics might suffice for options with a dominating alternative (Bettman et al., 1993). In summary, people tend to find tradeoffs more difficult than choices with a dominating alternative, which does not pose a decision conflict. Following the findings by Dhar and Simonson (2003), people may also be more inclined to defer these choices than choices containing a dominating alternative. In addition, we argued that difficult choices may be more likely to be delegated to AI. Thus, it follows:

H1. People are more likely to delegate decisions to AI in tradeoff condition 1 and 2, compared to decisions in the dominant condition.

3.2. Hypotheses for options that are equally good

As discussed in the previous section, findings are inconsistent as to whether options that are equally good are easier than options involving difficult trade-offs. On the one hand, a study by Willemsen et al. (2016), showed that a small diverse list was perceived as leading to less choice difficulty and enhanced choice satisfaction compared to a list with more similar options. On the other hand, the study by Kim et al. (2013) showed that the perceived similarity of two choice options decreased choice difficulty and choice avoidance. Their explanations were that small differences draw more attention to perceived similarity than identical features do, and that similarity may be based on the average of the most distinct features.

This explanation is related to the two different mental representations of perceived similarity discussed by Xu et al (2013). To summarize, one representation bases perceived similarity on how well options serve goal fulfillment. For people with this type of mental representation, increased similarities between options may decrease choice difficulty as two options are perceived as substitutes for reaching the same goal. This was also found by Kim et al. (2013), who reported that participants in the same price condition perceived the options as less substitutable.

The size of the choice sets does not explain why the results by Kim et al. (2013), were different from the study by Willemsen et al. (2016). The study by Kim et al. (2013) compared the perceived similarity between choices with six or fewer options. The study by Willemsen et al. (2016) found that increased diversity only decreased choice difficulty for sets with 5 items, which does not reflect a big difference in set size. One reason for the difference in results, however, may be the different presentations of choice attributes. Kim et al. (2013) provided participants with a numerical overview of the attribute values of each choice option. Willemsen et al. (2016) however, used a diversification algorithm based on latent features to diversify their choice lists. In practice, participants were presented with a list of five different movies and could not see how each movie scored on different characteristics. Therefore, participants in this study were probably less able to explicitly base their similarity judgements on common features. Thus, they might have been less inclined to see the options as substitutes, which might decrease choice difficulty.

As our research focuses on choices with a numerical representation of the two attribute values that are present, we expect that the findings by Kim et al. (2013) apply. These findings are also consistent with the inverted-U-shaped effect found by Scholten and Sherman (2006). Remember that they also found that the choice difficulty is lower for very small tradeoffs, compared to relatively large tradeoffs. Taken together, we only expect a small choice difficulty for choices that are equally good compared to a situation in which there is a relatively large trade-off. Hence, we predict:

H2. People will be less likely to delegate their decisions to AI in the equally good condition than options involving a relatively large trade-off (as in condition 4 and 5).

Studies on the choice between two alternatives have shown that equally attractive alternatives with few differences between the attributes seem to make choices more difficult compared to when one alternative is clearly superior (Dhar, 1997; Scholnick & Wing, 1988). To recap, the findings by Dhar (1977) suggest that equally attractive alternatives lead to indifference about which outcome is obtained, which increases choice deferral. It has been shown that this indifference may result in random choice (e.g. Gul et al., 2014). It is however also reasonable that people may delegate this decision to AI, as this process may still be more cognitively demanding. Therefore, we hypothesize:

H3. People will be more likely to delegate their decisions to AI in the equally good condition than in the dominant condition.

3.3. Hypotheses for options that are equally bad

The hypotheses so far were based on studies that all focused on options that were somewhat attractive. How does it work for sets containing options that are all unattractive? As discussed, research suggests that the control premium might not hold for delegation to AI (Candrian & Scherer, 2022; Bobadilla-Suarez et al., 2017). People may experience AI as less self-interested, and therefore, they may equally delegate decisions involving losses and gains to AI.

Although decisions between two equally bad choices does not involve a direct loss (e.g. a loss of money), reference dependence theory implies that they might be perceived as losses if people already have higher expectations of the type of hotel they would like to stay in. In our study, this may be likely as all attribute values in the equally bad choice sets are below a perceived neutral reference point. Because of this, people might have a prior expectation of a hotel that at least has average values on all its attributes. Still, as the control premium is originally proposed in the context of maximizing monetary payoffs, it is questionable whether such a "premium" also exists for the type of subjective choices we examine. It could still be that people delegate decisions involving equal subjective losses less than decisions involving gains to other people. Again, this might be because people are loss-averse and find it risky that others act on their behalf.

To our knowledge, only the study by Steffel and Williams (2017) might provide us with more insights. It has shown that overall participants are not more likely to delegate when two options are appealing than when they are both unappealing. In their research, participants faced an attractive and unattractive choice between two flavors of ice cream, which they had both given a highly similar bad or good ranking. Since this study was mainly hypothetical, it could still be that in reality, people are less willing to delegate choices between equally bad options to other people. Nevertheless, this might again not hold for delegation to AI, as it seems people experience them as less self-interested, and therefore seem to feel more in control. It is even less likely that people see AI as self-interested in the present research. This is because participants are made to believe the AI's choices are based on their personal preferences. Thus, we propose:

H4. If two options are equally bad, people will be equally likely to delegate their decisions to AI as when options are equally good.

Logically, it then follows that H2 and H3 also hold for options that are equally bad.

3.4. Delegation in the hotel vs. the monetary gambling context

As mentioned before, we look at whether our hypothesis holds for the two decision contexts we examine. The first one entails personal, hypothetical scenarios regarding the choice of a hotel. The second entails gambling decisions involving real monetary consequences. In summary, both contexts are similar in that the choice sets for these context similar characteristics as those described in Table 3. They are also similar in that participants, for both contexts, were told that the AI infers its choices on the participants' personal preferences.

Both conditions are also different in several ways, which may impact delegation rates differently. Firstly, people may feel less in control of their decision outcomes in the second context as they involve monetary gambles. If uncertainty would matter for the control premium and if an equivalent would exist for subjective choices, it follows that people would delegate less of their choices in the second context. According to the work by Owens, Grossman, and Fackler (2014) however, this aspect is not likely to create a difference in delegation between the two conditions.

Furthermore, it is possible that people are more inclined to gamble themselves as they perceive it to be enjoyable (e.g. Neighbors et al., 2002). This could lead to lower delegation rates in the second compared to the first condition. Finally, we are left with the possible effect of the stakes involved in both contexts. Even though the monetary gambling context involves real monetary consequences, it is unclear whether it has higher perceived states. It may just be as credible that people perceive the stakes to be higher in the hotel context, as they care more about choosing a hotel than receiving a small monetary bonus.

Taking these factors into account, we do not have a strong hypothesis about the differences in delegation rates between the monetary vs. the gambling context. Still, this study will explore whether there is a possible difference in delegation rates between the two contexts, while attempting to identify its underlying reasons.

4. Method

4.1. Basic experimental paradigm

In short, this study applies a 5-condition within-subjects experimental design. Our main independent variable of interest entails the type of decision situation or condition, in which: (1) one option dominates the other; (2) two options are equally bad; (3) two options are equally good. Moreover, we examine trade-offs in which; (1) one option has attribute values that are close together; (2) the attribute values of the two options are far apar. Apart from this, we measure some additional control variables that are described in Section 4.4.

Our main dependent variable includes people's decision to delegate a choice to AI, measured as a binary variable (yes or no). As mentioned before, we measure the effect of the independent variable for choices in two decision contexts, which are included as two separate parts of the experiment.

Each part of the experiment is carried out via lab.js, an online study builder (Henninger et al., 2020). In line with a within-subjects design, each participant faces all possible choices for all decision conditions and participates in both decision contexts. This design was chosen instead of a between-subject design, as it greatly reduces the required number of participants (also see Section 4.2.1).

4.1.1. Context 1: Choosing a hotel

In the first context, which includes the choice of a hotel, participants engage in 3 choice tasks per decision condition (which equals 15 choice tasks), in which they face a choice between two hotels that differ on two distinct attributes. These for instance include price and the time traveling to the city center. Before engaging in the choice tasks, however, participants first face 6 matching tasks[3]. These are all different than the choice tasks, but are based on the exchange rates between the different types of attribute pairs (e.g. price and time travelling to the city center), that were found in one of our pretests. These exchange rates were also used to construct the stimuli for the choice tasks in the equally good, bad and tradeoff conditions (see Section 4.3.1).

Because participants generally find it hard to fill out more negative values, participants either match on the most positive value (in terms of valence) of attribute A or B. This order is counterbalanced within-participants so that matching on attribute A is always followed by matching on attribute B. Additionally, the order of the matching tasks is randomized, to prevent order effects. The order of the choice tasks is randomized for the same reason.

In the experiment, participants are told that an AI is trained based on their responses in the matching tasks. In the choice tasks, participants can then choose to either delegate their choices to AI, or to choose for themselves. After participants complete their choices, they can immediately see the outcome of their (delegated) choices. In reality, no AI is trained, and choices by the AI are pre-programmed. More specifically, the AI chooses the dominant option in the dominant condition and chooses at random for the remaining conditions. These choices clearly do not reflect the participant's actual preferences. However, to increase the credibility of the AI, it was still important that people could immediately see the outcome of their choices.

[3] They also first participate in a point-allocation task, which we discuss into more detail in Section 4.4.1.

Figure 3.

Example of a matching task in the hotel context

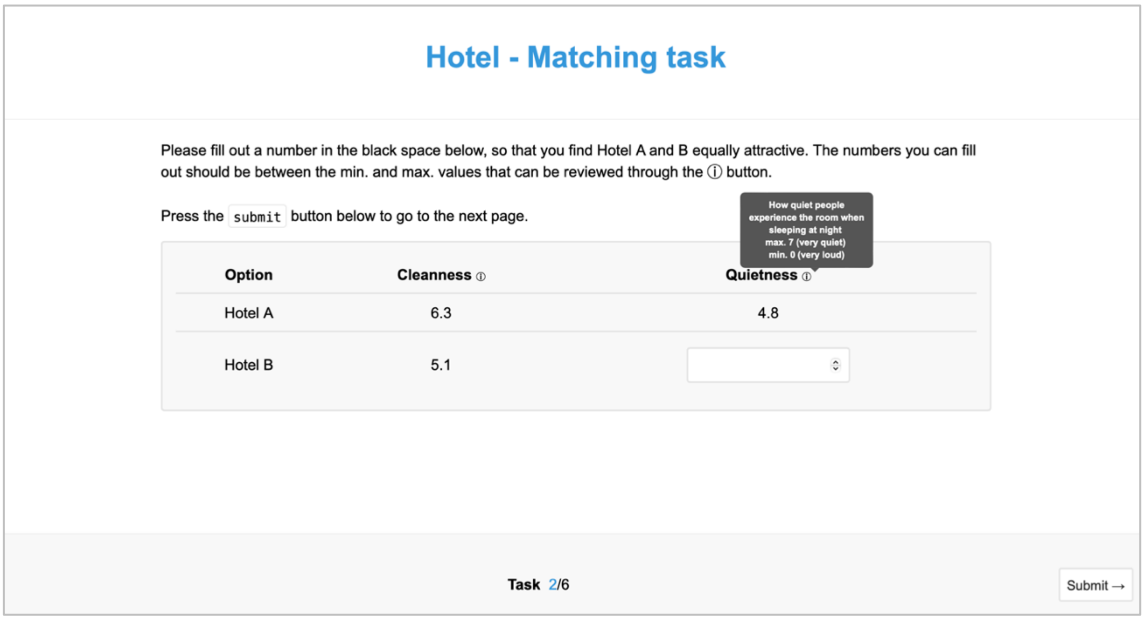
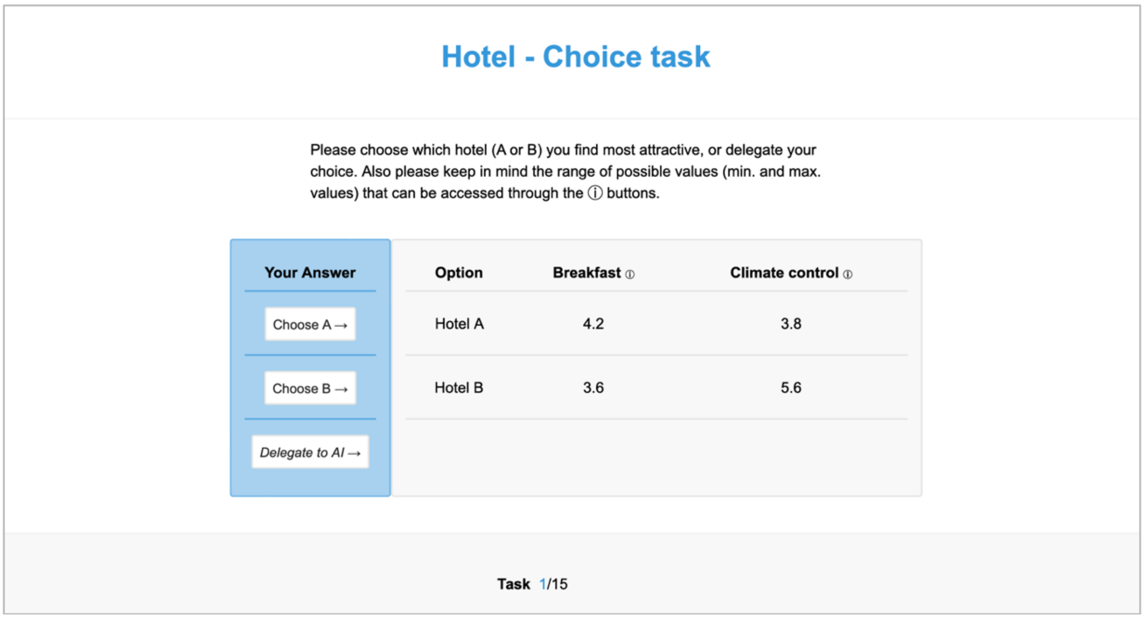


Figure 4.

Example of a choice task in the hotel context



4.1.2. Context 2: A monetary gamble

The second context, which includes a monetary gamble, has a highly similar design. However, participants in this context are asked to choose between two options that differ regarding a number of blue and red coins that can be gained or lost. Before making a choice, participants are told that after each separate choice, either the blue or red coin is worth twice as many points with a 50 percent probability.

Again, participants first face 6 matching tasks that are different than the choice tasks. For these tasks, we expected an exchange rate of 1, assuming equal importance of the blue and red coins (also see Section 4.3.2). The conditions in the matching and choice tasks are randomized in a similar way as in the hotel context. Moreover, participants are also told that an AI is trained based on their personal responses in the matching tasks. In the choice tasks, participants can then again choose to either delegate their choices to the AI, or to choose for themselves. Again, choices by the AI are pre-programmed and based on the dominant option or made randomly for the remaining conditions.

After completion of a choice task, participants can also immediately view the outcome of their choice. At the end of all choice tasks, participants again face the outcomes of the choices made by the AI and by themselves. The sum of points for these tasks is then translated into an additional monetary bonus, which was transferred to each participant (also see Section 4.2).

Figure 5.

Example of a matching task in the coin context

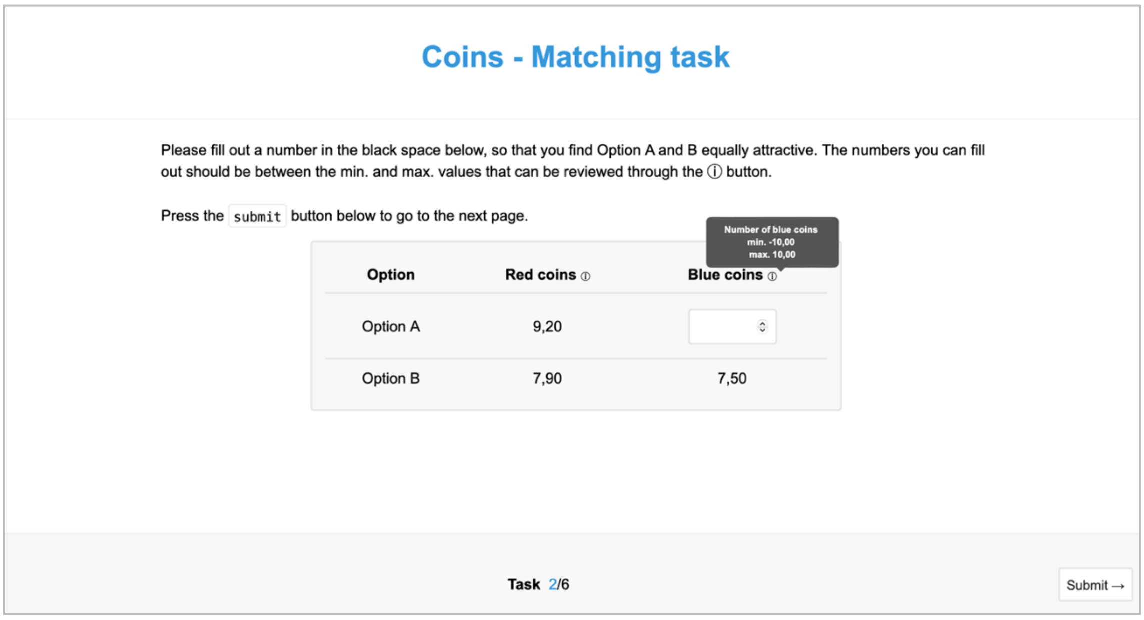
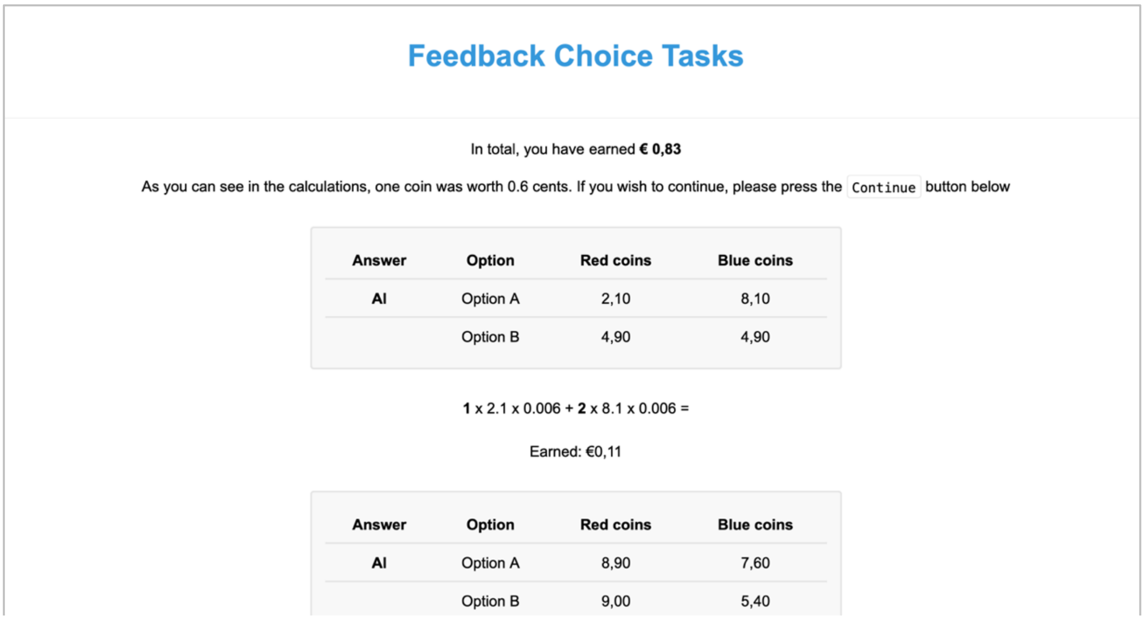


Figure 6.

Example of the monetary bonus that is calculated at the end of all choices



Feedback Choice Tasks

In total, you have earned **€ 0,83**

As you can see in the calculations, one coin was worth 0.6 cents. If you wish to continue, please press the `Continue` button below

Answer	Option	Red coins	Blue coins
AI	Option A	2,10	8,10
	Option B	4,90	4,90

1 x 2.1 x 0.006 + **2** x 8.1 x 0.006 =

Earned: €0,11

Answer	Option	Red coins	Blue coins
AI	Option A	8,90	7,60
	Option B	9,00	5,40

4.2. Participants

Participants for this study were recruited via the TU/e JFS Participant Database, plus via several WhatsApp groups. All participants (from within or outside of the TU/e) received a compensation of €5,00 for participating in the experiment. Additionally, they could earn a small bonus fee of up to €1,00 through effective gambling in the monetary gambling condition. Although people could also lose money in several gambling decisions, it was made sure that the additional bonus was always a positive amount of money.

Participants were excluded from selection for the study if they lacked normal (to corrected) eyesight, or if they could not read and/or comprehend the English language. In total, 96 participants participated. Generally, more younger people participated than older ones (*M* = 22, *SD* = 9.61, min. = 18, max. = 70). Moreover, most participants were female (n = 58), as opposed to male (n = 32), and people who did not prefer to specify their gender (n = 6).

Specifically of interest for the coin context, most participants were unoccupied (n = 40) or had a parttime job (n = 36), whereas a smaller subgroup had a full-time job (n = 20). In terms of income, most of them earned less than 1000 euros per month (n = 59)[4]. This seems logical as most of our participants indicated they were students (n = 66) and were unoccupied.

[4] This was followed by scale 2 (€ 1.000 - € 2.000 per month, n = 18), 3 (€ 2.000 - € 3.000 per month n = 10), 5 (> € 4.000 per month, n = 6) and 4 (€ 3.000 - € 4.000 per month, n = 3).

4.2.1. Sample size determination

To determine the appropriate sample size for a study, it is usually ideal to perform a power analysis based on the expected effect size. To our knowledge, there is no research that closely used our study procedure. Hence, we estimate an expected effect size based on studies that have looked at effect sizes for delegating difficult versus easy decisions to AI.

Existing studies on choice delegation to humans, found small to medium effect sizes for difficult versus easy choice decisions (also see Tversky and Shafir 1992, pp. 360-361). Steffel and Williams (2017, p.8) for instance examined whether participants would delegate difficult versus easy decisions regarding the choice of a new physician to their current physician. They found $d \approx 0.377$.

For our study, the smallest effect size of interest is the effect size between the equally good and the two moderate tradeoff conditions. We expect that the effect size between these conditions is smaller than the medium effect size found by earlier studies, as the difference between these conditions is more subtle. Particularly, we expected a 37% delegation-rate in the equally good condition and a 45% rate (based on the delegation rate found by Tversky and Shafir (1992)) in one of the tradeoff conditions. We think that an 8% delegation difference is still meaningful. Based on these means, we have run a power simulation in R with the *Superpower* package (Caldwell et al., 2022). From this, we found that a sample size of 120 would provide us with slightly more than 90% power, and hence we took this as our required sample size (see Appendix 7.1.1).[5]

4.3. Setting and stimulus materials

4.3.1. Context 1: Choosing a hotel

As mentioned before, participants chose between two hotels that differed in terms of two attributes. More specifically, the choice was always between two hotels that differed on a pair of characteristics (attribute A and B respectively), including (1) price and time travelling to the city center (2) cleanness and noise levels (3) breakfast and climate control. As an example, see Table 4 for an overview of the three choice tasks used for the dominant choice condition.

We chose the separate characteristics as they are inherently uncorrelated to each other. Furthermore, we chose the pairs in such a way that both characteristics in a pair are seen as equally important on average. This quality was verified through a pre-test, in which participants rated the importance of several hotel characteristics, through a point-allocation task. With this pre-test, we also verified whether the description and ranges of the hotel characteristics were complete and understandable. For a full description of these characteristics and their ranges, see Appendix 7.1.2.

[5] As we had 96 participants, we have not reached this number in practice.

Table 4.

Overview of the three choice tasks used for the dominant choice condition in the hotel context

Pair		**Attribute A**	**Attribute B**
1		Price	Time
	Hotel A	€ 90.00	45 min.
	Hotel B	€ 93.00	18 min.
2		Cleanness	Quietness
	Hotel A	6.1	4.5
	Hotel B	6.3	4.0
3		Breakfast	Climate control
	Hotel A	5.0	3.6
	Hotel B	3.5	3.8

We have drawn on expected utility theory (EUT) to compose the specific values for each attribute in the choice tasks. In short, we have simply calculated the expected utility of each hotel option through the following equation, based on Equation 3 (see Section 2.4.3, p.13):

$$E(u) = u(x_1) + u(x_2) = \sqrt{x_1} + \sqrt{x_2} \qquad (4)$$

where x_1 represents attribute A and x_2 attribute B.

Using this equation, we have composed the five choice sets for each pair of characteristics in such a way that the sets' average expected utility remained approximately equal for the dominant choice condition, the equally good condition, and the two trade-off conditions. For the first pair (price and time traveling to the city center), we for example have an average E(u) of 4.6 across both options for these conditions. For the equally bad condition however, the expected utility was naturally lower. Setting the expected utility within a pair for these conditions the same, we can exclude the possibility that higher delegation rates in for instance the dominant versus the equally good condition are explained by higher expected utility levels in one of the conditions. For a further elaboration on the construction and E(u) of these stimuli, see Appendix 7.2.

4.3.2. Context 2: A monetary gamble

In the gamble context, participants also faced three choice tasks for each choice condition. Similar to the hotel context, they thus faced 15 choice tasks. As described earlier, they chose between two monetary gambles that differed in terms of the amount of red and blue coins (attributes A and B respectively). See Table 5 for an example of choice tasks used for the different conditions.

For the stimuli used in the monetary gamble context, we have also drawn on EUT.
We can assume that both the number of red and blue coins are seen as equally important. This would be logical as they have equal expected payoffs. Based on EUT and our experimental design, we have used the following equation to calculate the expected payoff of an option:

$$E(u) = 0.5\sqrt{2 \cdot x_1 + x_2} + 0.5\sqrt{2 \cdot x_2 + x_1} \quad (5)$$

where x_1 represents attribute A and x_2 attribute B.

To avoid confounds, we have also composed the four choice sets for each pair of characteristics in such a way that the sets' average expected utility remained approximately equal for the dominant choice condition, the equally good condition, and the trade-off condition, again with the exception of the equally bad condition.

Table 5.

Overview of the four choice tasks used for the different conditions in the gamble context for E(u) = 2.81

Condition	Gamble	Red coins	Blue coins
Dominant	A	1,10	3,60
	B	0,80	5,20
Equally bad	A	-1,70	-3,80
	B	-1,80	-3,70
Equally good	A	3,10	2,20
	B	3,20	2,10
Tradeoff 1	A	0,30	5,30
	B	2,60	2,60
Tradeoff 2	A	1,40	3,90
	B	3,90	1,40

4.3.3. Pretest

To validate the stimuli we constructed according to EUT and MAUT and our general assumptions, we conducted a pre-test with 11 people that were familiar to the researcher. The purpose of this test was to see whether the two options in the equally good, bad and tradeoff conditions were deemed equally attractive. To answer our main question, participants faced the choice tasks that we had initially constructed for both the monetary gamble and hotel context. Prior to this, they faced matching tasks that were based on the values of the different choice tasks. Without going into too much detail, we finally based our new stimuli on the participants' mean and median matched values as they not always correspond to the matched values that we expected according to EUT. As these values did not drastically differ from our expected values, most expected utilities that were calculated through EUT and MAUT remained approximately the same (also see Appendix 7.2).

4.4. Measurements

As mentioned before, this study's main variable of interest is whether people decide to delegate a decision to AI or not, measured as a binary variable. We however also measure several control variables. Some variables were measured before or during the choices in one of the decision contexts, whereas others were measured after the decisions were made for both contexts.

4.4.1. Self-reported importance scores

When we composed our stimuli, we assumed that the characteristics in each pair (e.g., price and time) were deemed equally important. This this might however not hold for all participants, as some may for instance highly value price over time. Due to this, the prominence effect may occur. Again, this effect may make choices easier and may, according to our hypotheses, lead to less delegation to AI. As this may confound our results, we first try to account for this by directly measuring people's self-reported importance scores for each hotel characteristic. For this, we use a point-allocation task, which, to recap, entails dividing 100 points among the different hotel characteristics, according to their perceived importance. Participants only faced this task before engaging in the hotel context.

4.4.2. The absolute difference in exchange rate

Apart from equal importance of characteristics, we assumed *risk aversion* for both choices in the hotel and gambling context. However, not everyone may be risk averse to the same extent, leading to different values for which they think two options are equally (un)attractive. We cannot fully account for this through direct measures such as the point-allocation task. Therefore, we measure people's matched values before each decision context. Based on these, we calculate the absolute differences between participants' exchange rates and the expected exchange rates for each choice context and condition specifically. We then use it as a control variable in our statistical analysis[6].

4.4.3. Decision time

In conclusion we measure people's decision time (DT) for all choices in both contexts. This was measured from the point that they started making a choice till the point that they either delegated a choice or chose themselves. Although we had no prior hypothesis regarding the effect of DT, we included it as a control variable as we deemed it possible that higher decision times could signify an increased decision difficulty, which in turn could lead to increased delegation rates.

[6] In the hotel context, we for example expected an exchange rate of 2.40 for the price-time pair respectively. If someone would have an average exchange rate of 1.40, the absolute difference would be 0.60.

4.4.4. Measurements at the end of the experiment

Trait indecisiveness

As mentioned before, choice difficulty due to choice characteristics might increase indecisiveness, which may spur delegation to AI. However, people might also be more indecisive as part of their personality. We measure this possible confound through the Indecisiveness Scale (IS) by Frost and Shows (1993), considered to be the most valid measure of indecisiveness. This scale originally includes 15 items ("I try to put off making decisions") and measured on a five-point scale (1=strongly disagree, 5=strongly agree). As argued by Rassin (2007) this scale seems to measure both domain-specific indecision and indecisiveness. Hence, he proposed a shortened version (including 11 items) which seems more valid to measure trait indecisiveness. We hence use this version instead.

Income

In the monetary gamble context, participants might experience a monetary bonus as less relevant if they have higher income levels. Therefore, they could take the different choice tasks less seriously, or be more likely to delegate to AI because of lower stakes. To control for this, we measure participants' net monthly income through five scales[7]. Furthermore, we measure whether participants have a full or part-time occupation.

Gambling enjoyment

As mentioned in the related work section, increased gambling enjoyment can lead to a decreased delegation to AI. To control for this, we measure this through several items from the scale that was validated by Lloyd et al. (2010). In the scale respondents rate how frequently their gambling behavior involves 11 motivations (e.g. "To relieve boredom"), using a 4-point Likert scale with the options "never," "occasionally," "fairly often," and "very often". This scale is originally proposed to measure three central motivations for gambling: mood regulation, monetary incentives, and enjoyment. In our study, we only use the six survey items that loaded positively (with an item loading higher than 0.3) on the enjoyment factor in the original study.

Choice strategies

A binary measurement of choice delegation might not provide us with a full understanding of why people decide to delegate their choices. Hence, each participant in the study also received an open question asking about their motivation to delegate two random choice sets in each context: "Why did you decide to delegate or not to delegate?" If they chose themselves, they received a question about their choice strategy: "How did you choose yourself?" Accompanying each question, participants saw the designated choice set as well as information regarding what they or the AI chose on their behalf.

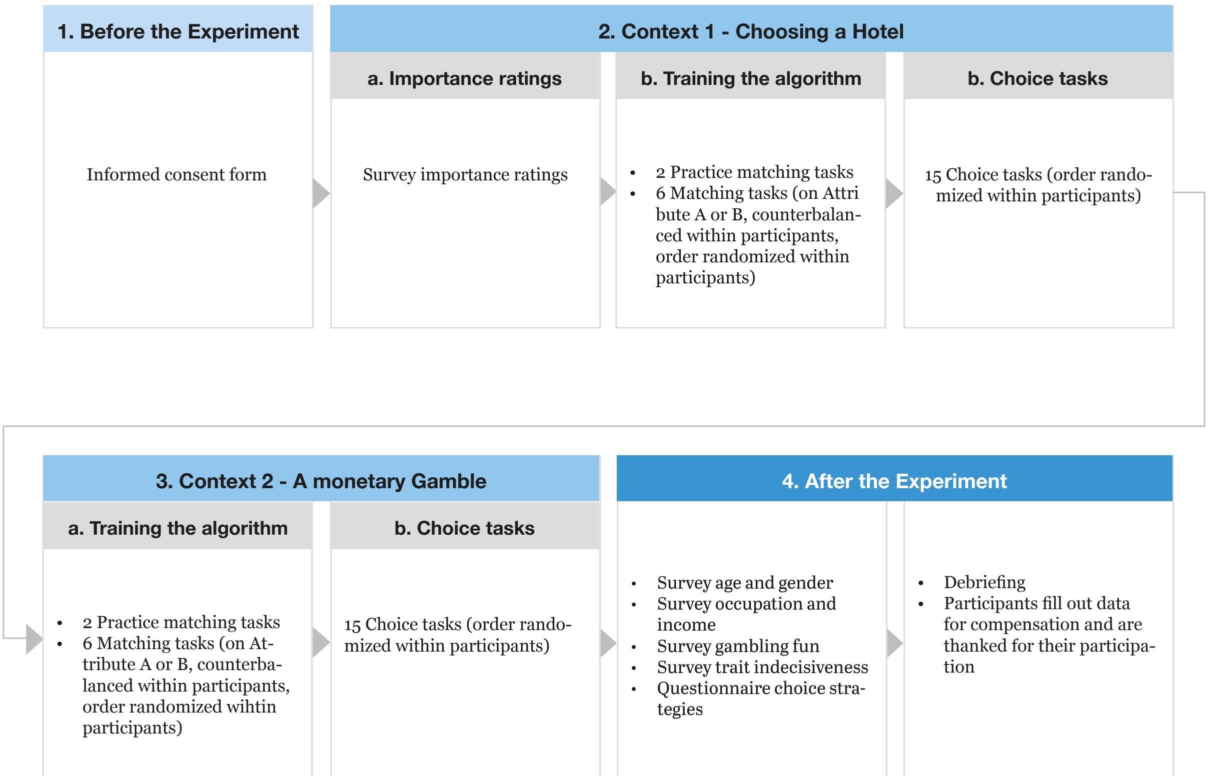
[7] 1 (<€ 1.000 per month), 2 (€ 1.000 - € 2.000 per month), 3 (€ 2.000 - € 3.000 per month), 4 (€ 3.000 - € 4.000 per month), 5 (> € 4.000 per month)

4.5. Procedure

As mentioned before, this study was conducted in an online environment. Before participation, participants were asked to read and sign an informed consent form that was also provided in this environment. Afterward, participants were introduced to the first part of the experiment, either entailing choices for the hotel or gambling context. In case they first engaged in the hotel context, they first read a brief explanation of the aim of that part of the experiment. This included descriptions of the different possible characteristics (and their ranges) of the different hotels and an explanation of how these could be reviewed during that experiment. Subsequently, they filled out a survey measuring participants' importance rates of the different characteristics. They were then informed about the matching task procedure.

Figure 7.

Procedure of the experiment



Note. In this example the hotel context is presented first, but the monetary gamble condition might also be first.

Before engaging in the 6 matching tasks, participants faced 2 practice matching tasks. In these practice matching tasks, participants were asked to match on a characteristic in a pair that was not present in the actual matching tasks (e.g. price-quietness). After each practice task, a pop-up would appear, showing the initial matching task, but including the participants' responses. Then, participants were instructed that they should find these options equally (un)attractive if they correctly completed their matching task. This feedback was useful to ensure participants' correct understanding of this task.

After the matching tasks, participants engaged in the 15 choice tasks. Then, they proceeded to the second part of the experiment, including the monetary gamble decisions. Up front, participants were informed about the possible bonus fee that could be earned from the different choice tasks. Other than in the first part, we did not ask participants to rate the importance of the two different types of coins, as they are logically equally important due to their equal expected payoffs. This was also confirmed by the pre-test results discussed previously.

Apart from that, the gambling part of the experiment followed a similar procedure as the hotel part. It also included 2 test- and 6 matching tasks, followed by 15 choice tasks. Following the experiment, participants first filled out a survey measuring gambling enjoyment. In addition, participants filled out a small questionnaire measuring trait indecisiveness, as well as two open questions regarding their strategies for their choices in the hotel and gambling contexts. After giving their consent, they first filled out a short questionnaire regarding their age and gender. Subsequently, they filled out a survey measuring their average income, and one assessing full-or part-time occupation. Finally, participants were debriefed and asked to provide some personal information to ensure payment of their compensation and acquired bonus fees.

4.6. Statistical Analysis

4.6.1. The effect of decision condition on delegation to AI

For our data analysis, we ran all quantitative models through STATA (StataCorp, 2019). For both the hotel and coin context, we evaluated our hypotheses through separate two-level logistic regressions, with a random intercept at the participant level. For these regressions, we include the decision condition (e.g. equally bad, dominant) as our main independent variable, combined with the different control variables for each context (e.g. gambling enjoyment). The dependent variable includes whether someone delegates a decision to AI. Regarding the control variables, the DT variable was first log transformed as the distributions of the observed DTs all seemed to be positively skewed for the conditions in each context. This means that a few large DTs or outliers would largely impact the found effects. Instead, log-transformed DT led to a normally distributed DT for each decision condition, which is generally regarded as an appropriate solution for decreasing the effect of outliers (Bland et al., 2013).

Before we ran our models, we conducted a manipulation check, to see whether the stimuli presented for each decision condition were indeed representative of that condition. As in our pre-test, we checked this by inspecting what percentage of participants would either choose option A or B. For example, an 80% choice of option A and a 20% choice of option B would indicate dominance, whereas we would expect 50% distributions for the other conditions. We conducted a second manipulation check to validate the exchange rates between the different pairs of attributes (e.g. price and time) that were found in the pre-test. Finally, we checked the different assumptions pertaining to a mixed logistic regression, such as linearity and no multicollinearity between the different independent variables.

4.6.2. Thematic analysis

To analyze people's choice strategies and motivation to delegate their decisions to AI, we conducted a small thematic analysis according to the method proposed by Braun and Clarke (2006). We separately conducted this analysis for both the hotel and coin contexts. Accordingly, we coded each text entry of a participant for both questions they received for each condition and subsequently grouped these entries in separate themes. Afterward, we examined which themes would most frequently occur for each decision condition.

5. Results

5.1. Indecisiveness and gambling enjoyment

For the survey scales of indecisiveness and gambling enjoyment, Cronbach's alpha was calculated. It was found that for both survey scales, all items have a relatively high internal consistency amongst each other (indecisiveness: all $\alpha >= 0.84$, gambling enjoyment: all $\alpha >= 0.76$). Overall, most participants' personalities leaned more toward the indecisive end of the spectrum ($M = 2.97$, $SD = 0.70$) and experienced low levels of gambling enjoyment ($M = 1.68$, $SD = 0.62$).

5.2. Results and discussion Context 1: choosing a hotel

5.2.1. Manipulation check

Differences in exchange rates

When looking at the average exchange rates based on the matched values for each pair (see Appendix 7.3.2), we find that the observed difference between the expected and observed exchange rates is biggest for the price-time pair ($M = 2.4$, $SD = 3.28$). This means that people seem to have quite diverse opinions regarding what amount of money they are willing to exchange for a minute less traveling time. As it is also furthest away from the exchange rate we expected, it seems that for this pair, one option might become more dominant to people. When examining the difference between the expected and observed exchange rates based on the point allocation task, we however find that this is largest for the cleanliness and quietness attribute pair ($M = 2.66$, $SD = 2.02$). In summary, the difference between the expected and observed exchange rates based on the matched values suggests that our stimuli might be least representative of conditions for the price-time pair. Based on the point-allocation task however, it seems they are least representative for the cleanness and quietness attribute pair.

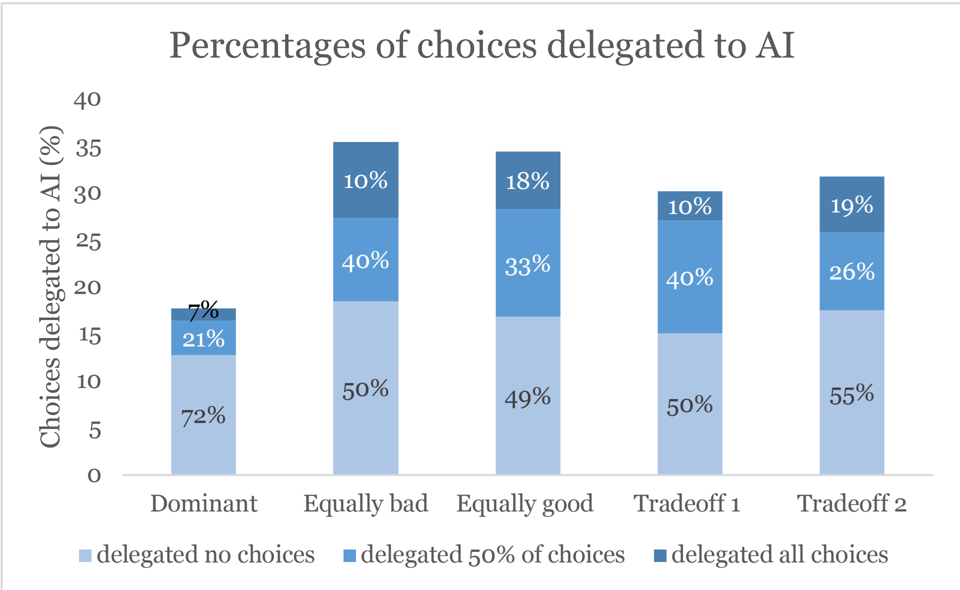
Validating the choice distributions

To gain a more concrete understanding of the representativeness of our stimuli, we look at the percentages at which options A and B are chosen for each condition of each pair (see Table 6-7, p. 33). Drawing from these results, we find that the stimuli for the price-time and breakfast-climate control pairs are quite representative. For instance, we find that for the dominant conditions, almost 90 percent chose option B for pair 3, and almost 100 percent chose option B for pair 1. For the other conditions, the values are also reasonably close to the expected 50 percent distributions. For the cleanness-quietness pair, however, we find some problematic values. This is because the stimuli for all conditions that should have 50 percent distributions for A and B seem to be representative of dominant conditions, and even seem to be more dominant than the original dominant condition. As this might confound our results we have excluded the data for this pair from our final analysis.

5.2.2. The effect of condition on the delegation to AI

Figure 8.

Percentage of choices delegated to AI for the hotel context



Note. This figure displays the percentage of choices that are delegated for each decision condition in the hotel context. Furthermore, the stacked percentages inside each bar represent the percentage of participants that have delegated either 0%, 50%, or 100% of their choices for that condition.

Before examining the main effect of decision condition on delegation, we first checked the no multicollinearity assumption between our control variables. This was met (see Appendix 7.3.3). When checking for linearity, however, we observed a quadratic relationship between log(DT) and delegation to AI. Hence, we also added the $\log^2$(DT) variable to our model. We also found rho=0.38, meaning that 38% of the variance in delegation is due to the variation at the participant level (or how frequently each individual participant delegates). This is also reflected in Figure 8, where one can see that even within each condition, there are large individual differences in delegation rates. Overall, this justifies our use of a multilevel regression instead of a regular one.

Tables 6, 7 and 8 provide an overview of the average percentages that are delegated for each condition for all pairs. At first glance, it appears for pair 1 and 3, decisions in the dominant conditions are delegated less than decisions in the other conditions (only 17% and 19%). There however does not seem to be a consistent difference regarding the equally bad and good conditions, and no difference between both the tradeoff and the other conditions (except for the equally good). In line with this, our multilevel logistic regression revealed a significant difference in delegation in the equally good condition versus the dominant condition (OR = 0.28, 95% CI [0.15, 0.49]). This means that people are 72% less likely to delegate a decision to AI in the dominant versus the equally good condition. In contrast, no significant differences were found between the equally good condition and the equally bad (OR = 1.06, 95% CI [0.64, 1.77]) and tradeoff conditions (OR = 0.76, 95% CI [0.45, 1.29] and OR = 0.73, 95% CI [0.44, 1.24] respectively). This confirms H4 but rejects H2. As the delegation rates in the dominant condition were significantly lower than in the other condition, we do find support for H1 and H3.

The average observed percentages of choices delegated for each condition can be found in Figure 8. The figure also demonstrates that for all conditions, most people do not seem to delegate a decision at all (almost 50% for each condition), followed by people delegating one choice per condition (21-40%), and people delegating all their choices (7-19%).

Table 6.

Percentages delegated and chosen for the stimuli of each condition for pair 1

	Price	**Time**	**Percentage chosen**	**Percentage delegated**
	Dominant			
Hotel A	€ 90.00	45 min.	1.25	16.67
Hotel B	€ 93.00	18 min.	98.75	
	Equally good			
Hotel A	€ 60.00	27 min.	50.72	28.12
Hotel B	€ 55.00	34 min.	49.28	
	Equally bad			
Hotel A	€ 195.00	73 min.	32.26	35.42
Hotel B	€ 199.00	67 min.	67.74	
	Tradeoff 1			
Hotel A	€ 83.00	26 min.	58.46	32.29
Hotel B	€ 97.00	17 min.	41.54	
	Tradeoff 2			
Hotel A	€ 75.00	29 min.	66.67	31.25
Hotel B	€ 103.00	11 min.	33.33	

Table 7.

Percentages delegated and chosen for the stimuli of each condition for pair 2 (was excluded from the data analysis)

	Cleanness	**Quietness**	**Percentage chosen**	**Percentage delegated**
	Dominant			
Hotel A	6.1	4.5	37.10	35.42
Hotel B	6.3	4.0	62.90	
	Equally good			
Hotel A	4.2	7.0	22.03	38.54
Hotel B	4.3	6.9	77.97	
	Equally bad			
Hotel A	0.8	2.0	14.04	40.62
Hotel B	0.9	1.9	85.96	
	Tradeoff 1			
Hotel A	6.0	5.1	82.43	22.92
Hotel B	5.5	5.6	17.57	
	Tradeoff 2			
Hotel A	6.3	4.8	83.12	19.79
Hotel B	4.8	6.3	16.88	

Table 8.

Percentages delegated and chosen for the stimuli of each condition for pair 3

	Breakfast	**Climate control**	**Percentage chosen**	**Percentage delegated**
	Dominant			
Hotel A	5.0	3.6	89.74	18.75
Hotel B	3.5	3.8	10.26	
	Equally good			
Hotel A	3.5	4.5	56.14	40.62
Hotel B	3.6	4.3	43.86	
	Equally bad			
Hotel A	3.1	1.5	40.32	35.42
Hotel B	3.0	1.7	59.68	
	Tradeoff 1			
Hotel A	4.4	3.6	41.18	29.17
Hotel B	4.0	4.8	58.82	
	Tradeoff 2			
Hotel A	4.7	3.5	48.48	31.25
Hotel B	3.5	7.0	51.52	

For the control variables, we found no significant effect of gender on delegation (e.g. for male with female as reference: OR = 1.06, 95% CI [0.64, 1.77]), and no significant effect of a difference in exchange rate based on either the point-allocation task (OR = 0.92, 95% CI [0.83, 1.02]), or participants' matched values (OR = 0.89, 95% CI [0.77, 1.02]). Furthermore, we did not find a significant effect of indecisiveness (OR = 1.26, 95% CI [0.74, 2.16]), age (OR = 1.03, 95% CI [0.99, 1.07]), and the number of occurrence of a choice trial (OR = 1.00, 95% CI [0.97, 1.03]). We did however find a negative effect of log(DT) on delegation to AI, with OR = $0.20*10^{-2}$ and 95% CI [0.00, 0.07]. Our regression also demonstrated a significant positive effect of $\log^2$(DT) on the delegation to AI (OR = 1.49, 95% CI [0.00, 1.24]). Interestingly, we also found a quadratic relationship between log(DT) and delegation to AI (also see Appendix 7.3.4). This suggests that for low DT's, people will delegate more to AI than for "average" DT's of around 8000 ms, whereas they will again delegate more for higher DTs.

5.2.3. Results of thematic analyses on qualitative data

Motivations to delegate

Table 9.

The different themes describing people's motivation to (not) delegate, illustrated by quotes

Theme	Quote
Trust in AI	"Because I was confident that the AI was making correct choices and the difference for me didn't seem that big." [p.74, a reason to delegate] "I value cleanness over quietness and I felt like the AI would not know this." [p.15] "I was curious if the AI had picked up on the fact that I want to decide for a cheaper hotel." [p.63]
Maintaining control	"I prefer to do things on my own." [p.20] "They were fairly the same and I prefer cleanness so I wanted to make sure that option was chosen." [p.18, a reason not to delegate]
Uncertainty about preferences	"It looked very similar so I could not choose." [p.12, a reason to delegate] "I prefer cleanliness over quietness, I choose myself because I know this preference." [p.95, a reason not to delegate]
Anticipated regret	"I delegated because the difference in features made the two hotels equally appealing. If I had made the decision myself, I would've regretted it later. That's why I wanted something else to make the decision for me." [p.73]
Unimportance	"Because I did not have a strong opinion myself, so if it doesn't matter to me, it is more easy to delegate the choice." [p.13]

Recall that we questioned people's motivation to delegate (or to not delegate) a decision to AI for two random choice sets in the hotel context. Specifically, we asked the question "Why did you decide to delegate or not to delegate?" Based on participants' answers, we have specified the 5 themes in Table 9. Taking these into account, it is likely that participants delegate less to AI in the dominant condition because of two reasons. Firstly, the uncertainty of preferences, or indifference, seems to be less present in the dominant condition. Secondly, it seems that participants in this condition prefer to remain more in control. For the other conditions, it is reasonable that unimportance plays a larger role in increasing delegation rates. Table 10 seems to support this hypothesis as we found more codes that were classified as *unimportance* in the equally good, bad, and tradeoff conditions. Besides, all codes related to uncertainty about preferences were found for these conditions as well. Surprisingly, there are also many participants who indicated to be certain of their choice in the large-tradeoff conditions. However, in these conditions participants seem to place less value on control. We will further discuss possible reasons for this difference in the Section 6.

Table 10.

The number of codes classified according to each theme per condition, describing people's motivation to (not) delegate

Condition	**Trust in AI**	**Maintaining control**	**Uncertainty about preferences**		**Anticipated regret**	**Unimp.**
			Certain	Uncertain		
Dominant	4	5	12	0	0	0
Equally good	9	0	10	4	0	1
Equally bad	3	1	3	9	0	2
Tradeoff 1	8	1	8	7	0	1
Tradeoff 2	5	0	13	3	1	2

Note. Unimportance has been abbreviated to Unimp.

5.2.4. Choice strategies

Table 11.

The different themes describing people's strategies when choosing for themselves, illustrated by quotes

Theme	**Quote**
Prominence	"I was pretty sure that I care more about climate control and not about breakfast." [p.11]
A balanced experience	"I preferred a more average experience but worried the AI might think I preferred extreme cleanliness." [p.30] "Since breakfast and climate control are equally important to me, I could just add up the scores and chose the higher one." [p.3]
Choosing randomly	"Difficult decision because I see not that much difference between the options. I pick one randomly." [p.86]

As mentioned before, participants received the follow up question "How did you choose yourself?" if they decided to make a choice of the two random choice sets themselves. Generally, we have classified participants' strategies for this according to three different themes: *Prominence, A balanced experience* and *Choosing randomly* (see Table 11).

Participants thus seem to apply different strategies when choosing by themselves. We also observe that participants may delegate because of indifference, but this might just as well be a reason for choosing by themselves.

Drawing from the numbers in Table 12, it could be possible that mainly going for a balanced experience makes it easier for people to justify their choice, decreasing choice difficulty and delegation rates. To illustrate, we see that many participants (n = 10) went for a balanced experience in the dominant condition, which is the condition with the lowest delegation rate. The number of codes for the prominence effect does not seem to correlate with lower delegation rates but does partly explain why some participants do not delegate their choices in the equally good and second tradeoff condition.

Table 12.

The number of codes classified according to each theme per condition, describing people's strategies when choosing for themselves

Condition	Prominence	A balanced experience	Choosing randomly
Dominant	4	10	0
Equally good	9	8	0
Equally bad	2	3	1
Tradeoff 1	2	3	0
Tradeoff 2	10	5	2

5.3. Results and discussion Context 1: a monetary gamble

5.3.1. Manipulation check

Difference in exchange rates

When looking at the average exchange rates based on the matched values for each pair (see Appendix 7.3.2), we found quite large differences between the expected exchange rate of 1/1 and the observed exchange rates for each utility level (min. = 0.77, max = 0.91). This difference did not seem to largely vary between utility levels. In sum, this suggests that the difference between the expected and observed exchange rates based on the matched values suggests that our stimuli might not entirely be representative of the different decision conditions.

Validating the choice distributions

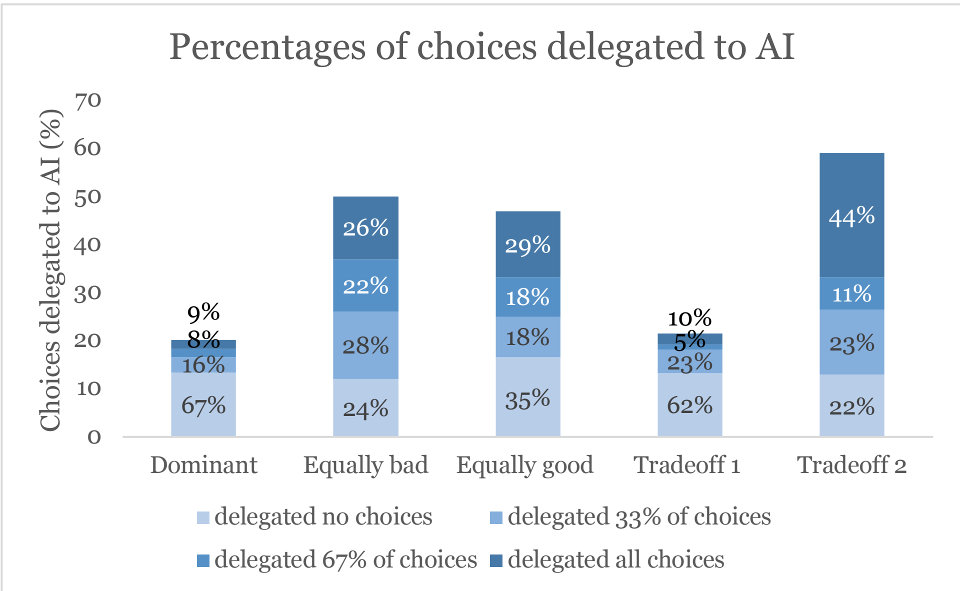
To gain further insights, we again also look at the percentages at which options A and B are chosen for each condition for each utility level (see Table 13-15, p. 39). Contrary to what we found above, these seem to confirm that the stimuli for all pairs are quite representative for all utility levels. As an example, we find that for the dominant conditions, the dominant option is chosen between 68 and 83% of the time.

One thing to note is that although most stimuli seem to be representative, they still do not perfectly represent each condition. For level E(u) = 2.81 for instance, option B is chosen 71% of the time for the equally bad condition. This seems to be more indicative of a dominant condition. One possible explanation is that some people look at which option is least negative on the minimum value that is present. For instance, they might prefer option B with -1,80 red and -3,70 blue coins instead of option A with-1.70 red and -3,80 blue coins as -3,70 is higher than the minimum value of -3,80. Nevertheless, this explanation does neither reflect principles from EUT, nor from prospect theory. Furthermore, this explanation does not account for the chosen percentages for the equally bad condition for E(u) = 3.81, which is close to the expected 50/50 distribution. Because we, therefore, lack a clear theoretical reason for excluding any data from our analysis, we have run our analysis with all data included.

5.3.2. The effect of condition on the delegation to AI

Figure 9.

Percentage of choices delegated to AI for the coin context



Note. This figure displays the percentage of choices that are delegated for each decision condition in the coin context. Furthermore, the stacked percentages inside each bar represent the percentage of participants that have delegated either 0%, 33%, 67% or 100% of their choices for that condition.

Before examining the main effect of decision condition on delegation, we first checked the no multicollinearity and linearity assumptions. Both of these assumptions were met (also see Appendix 7.3.3 for the correlations between variables). Additionally, we found rho=0.46, meaning that 46% of the variance in delegation is due to the variation at the participant level. Again, Figure 9 shows that individuals may largely differ with regards to delegation in each condition. Once more, this justifies our use of a multilevel regression instead of a regular one.

Again, Table 13, 14 and 15 display the delegation percentages for all conditions, this time for each utility level. Eventually, the logistic regression revealed a significant difference in delegation in the equally good condition versus the dominant condition (OR = 0.04, 95% CI [0.09, 0.23]), and the equally good versus the first (OR = 0.16, 95% CI [0.10, 0.26]) and second (OR = 1.88, 95% CI [1.24, 2.84]) tradeoff condition. This means that people are less likely to delegate a decision to AI in the dominant and first tradeoff condition versus the equally good condition, and more likely to delegate in the second tradeoff condition versus the equally good condition. Hence, this confirms H3. It also supports H1 only for the second tradeoff condition, and only partially supports H2. No significant difference was found between the equally good and bad condition (OR = 1.30, 95% CI [0.86, 1.96]) and the dominant and first tradeoff condition (OR = 1.12, 95% CI [0.67, 1.89]). This thus confirms H4 and means that H1 is partially rejected for the first tradeoff condition. The observed average percentages of choices delegated for each condition can be found in Figure 9.

The figure also shows that for the second tradeoff condition, 44% of the participants delegated 100% of their choices, which is quite high in comparison to the other conditions. Furthermore, we see that for each condition a higher percentage of participants delegate 100% of their choices instead of 67%, which could be because participants prefer to delegate consistently.

Table 13.

Percentages delegated and chosen for each condition for E(u) = 2.81

	Red coins	**Blue coins**	**Percentage chosen**	**Percentage delegated**
	Dominant			
Option A	1.10	3.60	32.43	22.92
Option B	0.80	5.20	67.57	
	Equally good			
Option A	3.10	2.20	59.62	45.83
Option B	3.20	2.10	40.38	
	Equally bad			
Option A	-1.70	-3.80	29.41	46.88
Option B	-1.80	-3.70	70.59	
	Tradeoff 1			
Option A	0.30	5.30	40.85	26.04
Option B	2.60	2.60	59.15	
	Tradeoff 2			
Option A	1.40	3.90	55.56	62.50
Option B	3.90	1.40	44.44	

Table 14.

Percentages delegated and chosen for each condition for E(u) = 3.81

	Red coins	Blue coins	Percentage chosen	Percentage delegated
	Dominant			
Option A	7.10	3.40	86.08	17.71
Option B	7.40	1.60	13.92	
	Equally good			
Option A	5.20	4.50	56.25	50.00
Option B	5.30	4.40	43.75	
	Equally bad			
Option A	-6.10	-3.60	53.33	53.12
Option B	-6.20	-3.50	46.67	
	Tradeoff 1			
Option A	2.10	8.10	30.26	20.83
Option B	4.90	4.90	69.74	
	Tradeoff 2			
Option A	5.70	4.00	39.47	60.42
Option B	4.00	5.70	60.53	

Table 15.

Percentages delegated and chosen in the official test for E(u) = 4.81

	Red coins	Blue coins	Percentage chosen	Percentage delegated
	Dominant			
Option A	8.90	7.60	81.82	19.79
Option B	9.00	5.40	18.18	
	Equally good			
Option A	7.50	7.90	39.62	44.79
Option B	7.60	7.80	60.38	
	Equally bad			
Option A	-7.90	-7.50	31.25	50.00
Option B	-7.80	-7.60	68.75	
	Tradeoff 1			
Option A	7.80	7.60	51.90	17.71
Option B	9.90	5.70	48.10	
	Tradeoff 2			
Option A	6.90	8.50	36.36	54.17
Option B	8.50	6.90	63.64	

For the control variables, we again found no significant effect of gender on delegation (e.g. for male with female as reference: OR = 1.61, 95% CI [0.69, 3.76]), and no significant effect of a difference in exchange rate based on participants' matched values (OR = 1.24, 95% CI [0.82, 1.87]). Plus, we did not find a significant effect of indecisiveness (OR = 0.90, 95% CI [0.54, 1.49]), income (all p> 0.05), and the number of occurrence of a choice trial (OR = 1.00, 95% CI [0.98, 1.03]). Yet, we found a negative effect of log(DT) on delegation to AI, with OR = 0.47 and 95% CI [0.37, 0.60]. Furthermore, we found that people who enjoyed gambling more also delegated less (OR = 0.45, 95% CI [0.23, 0.89]), and that older people tended to delegate more (OR = 1.08, 95% CI [1.02, 1.14]).

Notably, we also found a negative relationship between log(DT) and delegation to AI (also see Appendix 7.3.4). This suggests that for trials with longer decision times, people will delegate less to AI. In comparison to the hotel context, this might be because participants try to rationalize their choice through calculation, which leads to a higher decision time and less indecisiveness, but to a lower delegation probability. Apart from this relationship, we report a negative relationship of gambling enjoyment, indicating that people who enjoy gambling more are less likely to delegate their decision, which we also expected. Furthermore, we found a negative effect of age. In short, this could be because older people are less likely to be students at a technical university. Hence, they might not have understood the gambling task as well, resulting in indecisiveness or indifference, fueling choice delegation.

Overall, its seems that people tend to delegate more to AI in the second context compared to the first (40% vs. 30%, $p = 0.00$). As mentioned in Section 3.4, there might be several reasons for this difference, which we will elaborate on in the discussion.

5.3.3. Results of thematic analyses on qualitative data

Motivations to delegate

Table 16.

The different themes describing people's motivation to (not) delegate, illustrated by quotes

Theme	Quote
Trust in AI	"I thought the machine might be better at math than me." [p.45, a reason to delegate] "As this talk involved money, why it might be better to let an AI make the choice? I was afraid maybe the information I provided was not that good so I decided to risk it myself." [p.24] "Equal outcome and wondering about the pattern in the AI choice. A confidence interval would have been interesting." [p.9]
Maintaining control	"I choose myself because I was not ready to gamble and see what AI chooses for me. I wanted to have more control." [p.27]
Uncertainty about preferences	"Choice obvious for me, so do not let AI maybe make the wrong choice." [p.9, a reason to not delegate] "I could not clearly see which one was better so I let the AI choose." [p.18]

Anticipated regret	“I delegated because the difference in features made the two hotels equally appealing. If I had made the decision myself, I would've regretted it later. That's why I wanted something else to make the decision for me.” [p.73] “If I had lost a lot of money on this choice I would've been mad at myself for letting the AI choose such an important decision, so I preferred to do it myself.” [p.46]
Unimportance	“Not interested in gambling.” [p.13, reason to delegate]
Time efficiency	“Even if I think more I might choose wrong, AI just makes quick decision.” [p.90]

Again, for the people’s motivation to (not) delegate their decisions to AI, we have defined the 6 themes in Table 16. In contrast to the hotel context, we also found the additional theme of *Time efficiency.* Table 17 also gives us an indication of how many codes have been classified as each theme for each condition. Although not all participants have stated their choice strategy for each condition, this table suggests that many participants were certain of their choices for the dominant and first tradeoff condition. This tells us why participants in these conditions delegated less of their choices to AI. For the second tradeoff condition, we observe more codes that relate to indifference. Possibly, this explains why more participants have delegated their choice in this condition. Finally, the results do not reveal large differences between the number of codes for the equally good and bad conditions. This implies that for these conditions, people seem to delegate for the same reasons, leading to approximately equal delegation rates.

Table 17.

The number of codes classified according to each theme per condition, describing people’s motivation to (not) delegate

Condition	**Trust in AI**	**Maintaining control**	**Uncertainty about preference**		**Anticipated regret**	**Unimp.**	**Time**
			Certain	Uncertain			
Dominant	3	3	13	3	6	2	0
Equally good	4	3	3	17	2	4	1
Equally bad	2	3	0	17	2	2	1
Tradeoff 1	6	4	14	2	4	1	0
Tradeoff 2	6	0	2	14	0	11	0

Note. Unimportance and Time efficiency have been abbreviated to Unimp. and Time respectively.

5.3.4. Choice strategies

Table 18.

The different themes describing people's strategies when choosing for themselves, illustrated by quotes

Theme	Quote
Intuitive or mathematical reasoning	"Option A gives two large amounts, whereas Option B gives one slightly larger and one fairly smaller amount." [p.44] "I chose the one with the lowest negative number." [p.57]
Risk aversion	"As I don't like to test my luck too often, I picked the more "safe, balanced" choice which might not earn me a lot of money, but I wouldn't lose a lot of money either." [p.73]
Choosing randomly	"Both options are the same to me. So I just picked one randomly." [p.21]

Again, using the strategies described in Table 18 using may make it easier for people to justify their choice, which may lower delegation rates. As an example, from Table 19, we see that reasoning strategies seem to correlate with lower delegation rates in the dominant condition. Furthermore, risk aversion seems to affect the decreased delegation rates in the second tradeoff condition as well. As described above, this might have been because of the increased dominance of the option with values that closer together.

As in the hotel context, we also observe that unimportance can both be a reason for delegation and for choosing by yourself. To clarify, we have classified 8 codes as *Choosing randomly* for the second tradeoff condition, which is also the condition with the highest delegation rates.

Table 19.

The number of codes classified according to each theme per condition, describing people's strategies when choosing for themselves

Condition	Reasoning	Risk aversion	Choosing randomly
Dominant	16	7	0
Equally good	5	4	4
Equally bad	2	2	5
Tradeoff 1	12	14	3
Tradeoff 2	6	0	8

6. General discussion

This study was motivated by the presumption that the delegation of subjective choices to AI can provide several advantages to people such as time-effectiveness, and avoidance of indecisiveness, which might for instance lead to tunnel vision and choosing the default option (Rassin, 2007; Shafir et al., 1993). In general, we hypothesized that people may be more likely to delegate difficult decisions compared to easy ones and that this difficulty may depend on choice characteristics. In an online experiment, we have examined people's delegation rates to AI for two different contexts: one regarding the choice of a hotel and the other regarding the choice of a monetary gamble. For each context we also examined whether people's delegation rates differed for five different choice conditions, each with different characteristics.

The results from the hotel context showed that people are less likely to delegate decisions with a dominant option compared to decisions that either entail equally good or bad options or a tradeoff between options. In the monetary gamble context, we found different results; both the dominant and first tradeoff conditions resulted in lower delegation rates than the other conditions. Importantly, the difference in delegation rates for both contexts could not be attributed to personal perceptions of dominance as we controlled for people's absolute differences in exchange rates based on their matched values and point allocation. Furthermore, it could not be ascribed to differences in utility levels as these were kept equal across all conditions (except for the dominant one) for each utility level.

6.1. The equally good and bad vs. the tradeoff condition

Originally, we expected that if two options are equally good or bad, people will be less likely to delegate their decisions to AI than if options involve a relatively large trade-off. This was because we expected a small choice difficulty for the choices in the former conditions compared to a trade-off condition, since people may see the options of these choices as substitutes. Nevertheless, we found that for the hotel context, the delegation rates between both tradeoff conditions and the equally bad and good conditions did not significantly differ. If anything, our results imply a reverse pattern, namely that people delegated more in the equally bad and good conditions, although this difference was not significant.

Still, our findings might be consistent with the inverted-U-shaped effect found by Scholten and Sherman (2006). Initially, we viewed our tradeoff condition as a medium tradeoff, which according to the inversed-U-shaped effect, would lead to a higher perceived choice difficulty and higher delegation rates. Yet, participants may have perceived our tradeoffs as large tradeoffs, which might have decreased choice difficulty to the point that the choice was as easy as a choice for options that are equally bad or good. As found by Scholten and Sherman (2006) this effect may be due to argumentation: people find it easier to justify their choices for large tradeoffs.

Unfortunately, our qualitative data does not provide us with sufficient insights into this topic. Nonetheless, it does seem to suggest that the prominence effect was prevalent in the in the second tradeoff condition (as 10 codes related to this). This might indeed have made it easier for participants to justify their choices for this condition. Yet, the qualitative data does not provide us with a good reason for why the delegation rates in the first tradeoff condition were equal to those of the equally bad and good conditions. Hence, it seems less likely that argumentation plays a role for this type of tradeoff.

Another rationale for the small difference between the equally good and bad and both tradeoff conditions is that Scholten and Sherman (2006) found that when attributes have unequal importance, the inverted-U-shape becomes a “regular” U-shaped effect, meaning that intermediate tradeoffs are seen as less difficult than large and small tradeoffs. Their explanation is that when tradeoffs are small, people are concerned about the weak arguments in favor of the option with the highest value on their preferred attribute. However, if the difference is large, they worry about the sacrifices incurred by choosing their preferred option. As our results suggested unequal attribute importance all attribute pairs, this might explain why people found the choices in the tradeoff conditions relatively easy, and therefore delegated less of their choices for these conditions.

For the coin context, we only found that people are more likely to delegate in the second tradeoff condition than in the equally good condition. As discussed before, risk aversion seems to be the reason for the decreased delegation rates in the first tradeoff condition. These findings are partly in line with H2, meaning that some tradeoffs might be perceived as more difficult, increasing uncertainty and delegation to AI. Looking at the theory of Scholten and Sherman (2016) we suspect that in the gambling context, people find decisions including tradeoffs more difficult than those including equally good options. In this case, argumentation does not seem to decrease choice difficulty for large tradeoffs as both blue and red coins can be seen as equally important. Hence, people’s perceived choice difficulty may only depend on the sacrifices made when choosing one option over the other.

6.2. The equally good vs. the equally bad condition

Our fourth hypothesis stated that if two options are equally bad, people will be equally likely to delegate their decisions to AI when options are equally good. For both context, this was confirmed. Initially, we predicted this because findings on the control premium showed that it would be likely that this would hold for delegation to AI. This means that people are equally likely to delegate losses and gains to AI as they envision it to be less self-interested. However, we did not empirically test whether the effect of the control premium indeed explains this difference.

Furthermore, based on the observed delegation rates, people seemed to delegate slightly more in the equally bad vs good conditions for both contexts, although this effect was not significant. A reason for this might be that for equally bad decisions, people feel more anticipated regret. This was partly supported by our qualitative findings and by the study by Steffel and Williams (2017). They found that for equally bad options, people would be more likely to have anticipated regret and to opt out of decisions. Furthermore, they found that delegation could be an appealing alternative to opting out. Given that H4 was rejected in both contexts, it seems that this possible effect of anticipated regret is at least not large enough to cause a significant difference between these conditions.

It could still be that the effect of the control premium has partly canceled out the effect of anticipated regret for equally bad choices. That is, some people might have delegated more because of anticipated regret, whereas others might have delegated less because they wanted to feel more in control over decisions that involve equally bad versus equally good options. Consequently, more research is necessary to confirm both effects.

To measure the effect of the control premium, one suggestion would be to use a similar procedure as Candrian and Scherer (2022). To recap, they compared differences in delegation rates to humans vs. AI and how they would change when a decision outcome would either entail a monetary loss or gain. They also measured the perceived intentional capacity of both humans and AI. Similarly, our study could be extended by measuring these aspects for decisions in both the equally bad and good conditions. Lastly, to measure anticipated regret, one could for example adopt the survey scale developed by Marcatto and Ferrante (2008).

6.3. Overall delegation rates in the coin vs. the hotel context

Comparing the overall delegation rates between the hotel and coin context, we found that these were higher in the coin context. This might be due to two reasons. Firstly, although participants were told that the AI in both the coin and hotel context based its choices on their preferences, they could still have perceived the coin context as a more objective or “AI-appropriate” domain. This might especially be true as research has shown that mainly perceived (and not the actual) objectivity of a task determines people’s trust in the delegation to AI (Castelo et al., 2019). Finally, people might have had higher perceived stakes in the hotel context. To clarify, they may perceive a successful hotel stay as more important than a small monetary bonus. As mentioned before, this might decrease delegation rates (Ashoori & Weitz, 2019). Although both reasons seem likely, a future replication study would benefit from empirically investigating people’s perceived stakes and trust in AI for the two different contexts.

6.4. The effect of log(DT)

As mentioned before, in the hotel context people delegated less decisions to AI for more average decision times than for lower decision times, whereas they again delegated more frequently for higher decision times. This could probably reflect three different evaluation and decision strategies. Firstly, people might quickly scan the two hotel options, and decide to delegate because they are indifferent between the options, without a careful evaluation of the attribute values. Secondly, some people might carefully tradeoff the different attribute values against each other, which might help to justify their choice, reducing delegation. Thirdly, people may delegate to avoid further cognitive effort correlated with longer decision times. For the coin context however, we only found that delegation rates were lower for higher decision times. This could indicate that if people have higher decision times, they might make a probability calculation, which might make it easier to justify their choices and may therefore decrease their delegation rates.

Regrettably, we can only speculate about these strategies as we could not differentiate between the time taken for evaluating choice options, and the time for deciding upon delegation. Still, it might be difficult to empirically differentiate between these decision times, as the two decisions might be taken in parallel[8], and both processes are hard to measure through self-reports and behavioral measures.

[8] Research for example, provides evidence that people may have parallel processes in the brain for both the evaluation of decision options and for making decisions about the actions that should be made to obtain the

Potentially, future studies could employ eye-tracking to measure the DT for both processes separately. Then, the DT could be derived from the time participants would either fixate on the different choice options or on the delegation menu. This might be appropriate as the decision-making literature has widely recognized eye-tracking for assessing choice behavior (e.g. Vass et al., 2018). As mentioned above, this method should be carefully considered as parallel processes in the brain might prevent the separate measurement of these processes.

6.5. Limitations

The current study has several limitations. First, it could be that the numerical presentation of our stimuli has affected our results. As discussed in the related work section, two-option two-attribute choices with a numerical representation might make it easier for people to base their similarity judgments on common features, and to see options as substitutes (Xu et al., 2013). In the present study, this may have increased people's indifference which may increase delegation to AI. Yet, decisions in the real world may also involve choices between options with latent or "hidden" features. In that case people may be less inclined to base their judgments on common features and may therefore experience an increased decision difficulty. On top of that, choices may involve more than two attributes or options. Research has shown that comparisons between more options may be more difficult (e.g. Scheibehenne et al., 2010). Given this finding, people might for instance find choices between more equally bad or good options more difficult than in the current study. A strong recommendation for future research is to therefore to study the role of both latent features and the number of options on the delegation to AI.

Second, we do not yet understand people's mental models of the AI that was described in this study. As found form our qualitative data, some participants based their decision to delegate on their desire to test the AI's underlying mechanisms and competency. Others, however, chose not to delegate their decision because they did not believe that the AI would accurately represent their preferences. To further inspect the effect of mental models on people's delegation rates, future work could conduct this study with a real AI, while adding transparency about its competency.

A final limitation is that we were not able to adapt our stimuli based on people's personal preferences. To summarize, all participants in our study received the same stimuli for all conditions and we assumed that they would be indifferent between options in all conditions except the dominant one. We aimed to control for this difference though the absolute differences in exchange rates, but these were only based on two measurements. Therefore, participants different opinions regarding attribute weights may still have increased the variance in our data. To eliminate this effect, one recommendation for future studies would be to study the applicability of different methods for measuring preference elicitation, such as a conjoint analysis. As discussed by Louviere and Islam (2008), conjoint analysis is based on actual choice behavior and may therefore better predict attribute weights than matching tasks. Furthermore, one could consider training a latent-feature diversification algorithm (based on matrix factorization) such as the one used by Willemsen et al. (2016) to predict and construct choice sets with two options that people envision to be equally (un)attractive.

---

rewards related to these options (Rushworth et al., 2012). In this case, whether to delegate a decision could be seen as an action towards reward maximization.

6.6. Future research

Apart from the directions for future research that we suggested above, we envision two more. In the first place, we did not compare participants' delegation rates to AI against delegation to other people in the different conditions. Studying this is key, as research suggests that people may trust other people more than AI for subjective decisions (Castelo et al., 2019). Other research, however, found that people are more willing to take advice from AI over other people for both subjective and objective predictions (Logg et al., 2019). As these studies mainly focused on trust in AI, it is interesting to see whether this is also the case for a full delegation of subjective decisions. As mentioned before, examining the difference in delegation rates to humans vs. AI might also provide us with insights into the effect of the control premium. In conclusion, we were unable to determine whether participants would be equally likely to randomize their choice as they would delegate to AI. This might be the case since our results emphasize that indifference is a reason for delegation to AI. If this would be true, businesses and developers could save themselves the effort of implementing AI-driven delegation software.

6.7 Practical implications

Despite this study's limitations, it provides insights into the effect of choice characteristics on the delegation of subjective choice to AI. Most importantly, choice difficulty seems to play a role in whether people decide to delegate their decisions to AI. Furthermore, our qualitative results, despite being limited, suggest that personal differences seem to moderate people's perceived choice difficulty. For instance, we found that our participants may differ with respect to the need for control, indifference, anticipated regret, and levels of risk aversion. At last, our qualitative findings seem to confirm findings by prior research on the role of trust in AI in decision delegation.

Even though these results could be applied to the future development of AI-driven delegation software, this is more difficult than it seems. Because people may have different perceptions regarding what constitutes an easy decision for the different decision conditions we examined, algorithms should be able to make individual predictions on choice difficulty. Our results suggest that this prediction may be based on the type of decision condition, although these effects may differ for choices that concern personal preferences about gambling or purchases. Nevertheless, there are still other factors that may play a vital role in delegation, such as people's trust in AI and the effect of the control premium.

Having a thorough understanding of choice difficulty, however, can already inform businesses about the merits of investing in decision support for the type of decisions that are likely to be found difficult by a reasonable percentage of customers (which may vary for different businesses). As noted above, it is still complex to predict whether someone is willing to delegate a decision in a certain condition, as people's willingness to delegate a decision heavily relies on personal characteristics (and perhaps contextual factors such as time pressure). For customers, however, receiving the option to delegate a purchase decision in most appropriate cases might already give them the benefit of reduced cognitive effort.

Customers should however also know that companies could use delegation systems to manipulate them into making unfavorable decisions. Companies could for example use AI to recommend customers a small set of similar products, while telling them that based on their preferences, they would usually be indifferent between them. In line with our findings, people could then be more inclined to delegate this decision to AI. If the company's AI then not only takes the customers' need, but also profits into account, users may be deceived into buying more expensive products. Hence, it is at least necessary to be transparent about how accurately an AI represents an individual's preferences, as well as about its underlying mechanisms.

6.8. Conclusion

To our knowledge, our study provides the first important insights in the effect of choice characteristics on the delegation of subjective choices to AI. In this study, we have examined whether delegation rates to AI differ for five different choice conditions, each with its own characteristics. We measured this through an online experiment for two different contexts: one entailing the choice of a hotel and one the choice of monetary gamble. Our results suggest that choice difficulty, and the ability to justify one's choices, plays a large role in this delegation. In turn, this may depend on the characteristics of a choice. More specifically, this study showed that for both contexts, people are more willing to delegate decisions with a dominating alternative to AI compared to decisions including tradeoffs and equally (un)attractive options with small tradeoffs. Therefore, this seems to be a difference that might generalize for different types of subjective decisions.

Possibly, future research could identify more differences if the limitations of the present study are addressed. Furthermore, is still necessary to develop a method for constructing choice sets based on personal preferences. Another strong recommendation for future research is to study the role of both latent features and the size of a choice set on the delegation to AI. Findings on these topics can further guide developers of AI delegation systems in creating systems that are able to provide delegation opportunities to the right customers at the right time.

7. Appendix

7.1. Method

7.1.1. Power analysis in R

Particularly, we expected a 37% delegation-rate in the equally good condition and a 45% rate (in line with the delegation rate found by Tversky and Shafir (1992)). Based on these means, we have created the following discrete distributions, that we used for a power simulation in R with the *Superpower* package (Caldwell et al., 2022). Doing a power simulation in R we find the statistics written in Table 22. As we see, n = 120 gives us slightly more than 90% power, and hence we use take this as the required sample size.

Table 20.

Distribution in the tradeoff condition

Percentage delegated (%)	**0**	**16.7**	**33.3**	**50**	**66.6**	**83.3**	**100**
x	0	1/6	2/6	3/6	4/6	5/6	1
Pr(x)	0.10	0.12	0.24	0.26	0.14	0.10	0.04

Note. Condition has 6 choice sets, each choice can either be delegated or not. M_x=0.45 *SD*=0.26.

Table 21.

Distribution in the equally good condition

Percentage delegated (%)	**0**	**33.3**	**66.6**	**100**
x	0	1/3	2/3	1
Pr(x)	0.15	0.64	0.16	0.05

Note. Condition has 3 choice sets, each choice can either be delegated or not. M_x=0.37, *SD*=0.24.

Table 22.

Power effect size and required sample size for 5000 simulation runs

Amount of simulation runs	**Power**	**Effect size (Cohen's *d*)**	**Required n (sample size)**
5000	88.86%	0.32	100
5000	93.32%	0.32	120

Figure 10.

Example power analysis in R for n=100

```
>
> design_result <- ANOVA_design(design = "2w", n=120, mu = c(0.37,0.45), sd = c(0.24,0.26), r = 0.5, labelnames = c("COND", "good", "tradeoff"), plot = TRUE)
> simulation_result <- ANOVA_power(design_result, alpha_level = alpha_level, nsims = nsims, verbose = FALSE)
> simulation_result
Power and Effect sizes for ANOVA tests
           power effect_size
anova_COND 93.32     0.09906

Power and Effect sizes for pairwise comparisons (t-tests)
                          power effect_size
p_COND_good_COND_tradeoff 93.32      0.3217
```

Note. We set the standard deviations and means as specified above and use a within-subject design with two levels. We find effect size $d = 0.3217$ and power = 88.86%.

7.1.2. Descriptions and ranges of the different hotel attributes

Table 23.

Explanation of the different hotel characteristics

Attribute	Description	Range
Distance to the city center	Mean absolute distance from the hotel to the city center in km.	4.75 km (very far) - 0.00 km (very near)
Cleanness	How people experience the cleanness of the hotel room. For instance, is the room dusty, or is there mold in the shower?	0 (very dirty) - 7 (very clean)
Price	Price per night per room.	€45 (very cheap) - €300 (very expensive)
Quietness	How quiet people experience the room when sleeping at night.	0 (very loud) – 7 (very quiet)
Climate control	The extent to which people think that they are in control of the climate in their room. For example, is the room at its preferred temperature? Can the room be properly ventilated?	0 (very poor control) - 7(very good control)
Breakfast	How people experience breakfast. This includes opinions on the variety of drinks (e.g. coffee, tea) and food (e.g. different types of spreads or buns). It also includes opinions on taste and presentation.	0 (very bad) - 7 (very good)

7.2. The determination of E(u) in the different choice conditions

7.2.1. Hotel context

As mentioned before, we used Equation 4 (see Section 4.3.1) to have compose the choice sets for each pair of characteristics in the hotel context in such a way that the sets' average expected utility remained approximately equal for the dominant choice condition, the equally good condition, and the trade-off conditions (see Table 24).

As the first pair, price and time traveling to the city center, has characteristics that have different values ranges (0 to 80 min. and €45 to €285) compared to the other pairs (with ranges 0.0 to 7.0), its values were first normalized to a scale of 0 to 7 before calculating the E(u) of each option with Equation 4. For the normalization, we used Equation 6.

$$x_{normalized} = \left(\frac{x_{\max - x}}{x_{range}}\right) \cdot n \tag{6}$$

Here, $x_{normalized}$ is the normalized value, x_{max} the upper limit of the possible range of the original value, x the original value, x_{range} the range of the original value, and n the upper limit of the normalized value. As an example, we can normalize a given value for the travelling time attribute by calculating:

$$x_{normalized} = \left(\frac{80-x}{80}\right) \cdot 7 \tag{7}$$

For the equally bad condition, we have determined the average E(u) of a pair's choice set by subtracting the average E(u) of the sets in the other conditions from 7. For example, we have set the average utility of the choice set of pair 1 at 2.4 for the equally bad condition, which is calculated by subtracting 4.6 (the average utility for this pair in the other conditions) from 7. Finally, to control for differences in choice dominance, we have set the difference in E(u) between options the same ($E(u)_{dif} = 0.3$).

Note that with our approach we have some variation in the E(u) in each condition, as each choice set has a unique average E(u). We can however not interpret these E(u)'s as linear across pairs. For example, pair 1 with E(u) of 3.0 in the equally bad does not have an E(u) that is $\left(\frac{3}{2.3}\right) \approx 1.3$ times as high as pair 2. This comparison would not make sense as we do not have a measure of how relatively important each pair is compared to the other.

7.2.2. Coin context

For the equally bad condition in the coin context, we have determined the average E(u) of a pair's choice set by making negative the average E(u) of the sets in the other conditions. As in the hotel context, we made sure that for each dominant choice set, the difference in E(u) between the options remains equal ($E(u)_{dif} = 0.32$). As before, we also have some variation in the E(u) in each condition, as each choice set has a unique average E(u). Here, however, we can interpret these E(u)'s as linear across pairs, as each pair includes the same two attributes.

Table 24.

Average expected utility for each option per pair in each condition for the hotel and monetary gamble context

			Context 1 Choosing a hotel	**Context 2 Monetary gamble**
Condition	Pair	Option	E(u)	E(u)
Dominant	1	A	4.4	2.65
		B	4.7	2.97
	2	A	4.8	3.97
		B	4.5	3.65
	3	A	4.1	4.97
		B	3.8	4.65
Equally bad	1	A	2.4	-2.81
		B	2.6	-2.81
	2	A	2.3	-3.81
		B	2.3	-3.81
	3	A	3.0	-4.81
		B	3.0	-4.81
Equally good	1	A	4.7	2.81
		B	4.6	2.81
	2	A	4.7	3.81
		B	4.7	3.81
	3	A	4.0	4.81
		B	4.0	4.81
Tradeoff 1	1	A	4.6	2.85
		B	4.7	2.80
	2	A	4.7	3.90

		B	4.7	3.81
	3	A	4.0	4.81
		B	4.1	4.83
Tradeoff 2	1	A	4.6	2.81
		B	4.8	2.81
	2	A	4.7	3.81
		B	4.7	3.81
	3	A	4.0	4.81
		B	4.2	4.81

Note: The utilities are calculated for the final stimuli in the experiment, and thus were calculated for the stimuli based on participant's mean and median matched values in one of our pre-tests.

7.3. Results

7.3.1. Median and mean matched values

Table 25.

Hotel context: Mean and median matched values for each condition for all pairs

	Attr. A	Attr. B	$v_{matched}$(M)	$v_{matched}$ (Mdn)	Attr. $_{matched}$
	Price	Time			
Hotel A	92.00	29 min.	60.84	60	A
Hotel B	69.00	44 min.	39.40	40	B
	Price	Time			
Hotel A	65.00	31 min.	65.22	60	A
Hotel B	83.00	19 min.	26.62	27	B
	Cleanness	Quietness			
Hotel A	5.0	6.1	5.68	6	A
Hotel B	5.8	5.3	5.73	6	B
	Cleanness	Quietness			
Hotel A	6.3	4.8	6.02	6	A
Hotel B	5.1	6.0	5.63	6	B
	Breakfast	Climate control			
Hotel A	4.2	3.8	4.79	5	A
Hotel B	3.6	5.6	4.46	4.5	B
	Breakfast	Climate control			
Hotel A	4.5	3.5	5.48	6	A
Hotel B	3.5	6.5	4.60	4.5	B

Table 26.

Coin context: Mean and median matched values for each condition for all pairs

	Attr. A	Attr. B	$v_{matched}$(M)	$v_{matched}$ (Mdn)	Attr. $_{matched}$
	Blue coins	Red coins			
Option A	2,40	2,90	1.76	1.1	A
Option B	1,00	4,30	3.00	2.9	B
Option A	1,90	3,30	2.04	1.9	A
Option B	3,50	1,70	2.58	1.7	B
Option A	3,10	6,60	3.79	3	A
Option B	4,50	4,50	5.09	5.2	B
Option A	5,90	3,80	4.72	4.7	A
Option B	4,70	5,00	4.43	4	B
Option A	8,10	7,30	6.98	6.7	A
Option B	6,60	8,80	6.95	7	B
Option A	9,20	6,20	7.29	7.9	A
Option B	7,90	7,50	6.60	6.2	B

7.3.2. The difference in expected exchange rates

Table 27.

Hotel context: mean (absolute) differences in exchange rate for the different pairs

Pair	Diff. in exchange rate (mv)		
	M	Mdn	SD
Price-Time	2.40	1.30	3.28
Cleanness-Quietness	0.57	0.31	0.83
Breakfast-Climate control	0.60	0.51	0.47

Table 28.

Coin context: mean (absolute) differences in exchange rate for the different utility levels

E(u) level	Diff. in exchange rate (mv)		
	M	Mdn	SD
2.81	0.91	0.97	0.40
3.81	0.77	0.75	0.44
4.81	0.80	0.86	0.26

7.3.3. Correlations between control variables

Table 29.

Hotel context: Correlations between the different control variables

	Indecisiveness	**Age**	**DT**	**Log(DT)**	**Count trial number**	**Diff. in exchange rate (point all.)**	**Diff. in exchange rate (mv)**
Indeciveness	-	-.18	-.06	-.19*	.00	.02	-.14*
Age	-.18	-	.06	.19*	.00	-.02	.14*
DT	-.06	.06	-	.77*	-.15*	-.01	.04
Log(DT)	-.19*	.19*	.77*	-	-.26*	.02	.06
Count trial number (=place of occurance 1^{st}, 3^{rd} etc.)	.00	.00	-.15*	-.26*	-	.04	.01
Difference in exchange rate (point all.)	.02	-.02	-.01	.02	.04	-	.14
Difference in exchange rate (based on matched values)	-.14*	.14*	.04	.06	.01	.14	-

Table 30.

Coin context: Correlations between the different control variable.

	Gambling enjoyment	**Indecisiveness**	**Age**	**DT**	**Log(DT)**	**Count trial number**	**Diff. in exchange rate (mv)**
Gambling enjoyment	-	-0.24*	-0.18	-0.18*	-0.31*	0	-.10*
Indeciveness	-0.24*	-	-0.18	-0.18*	-0.31*	0	-.10*
Age	-0.18	-0.18	-	0.18*	0.31*	0	.10*
DT	-0.18*	-0.18*	0.18*	-	0.70*	-0.19*	-.06
Log(DT)	-.3103941*		0.31*		-	-0.19*	-.03
Count trial number (=place of occurance 1^{st}, 3^{rd} etc.)	0	0	0	-0.19*	-0.19*	-	-.05
Diff. in exchange rate (based on matched values)	-0.10*	-0.10*	.10*	-.06	-.03	-.05	-

7.3.4. The effect of log(DT) on delegation

Figure 11.

The effect of log (DT) on the predicted probability of delegating to AI for the hotel context

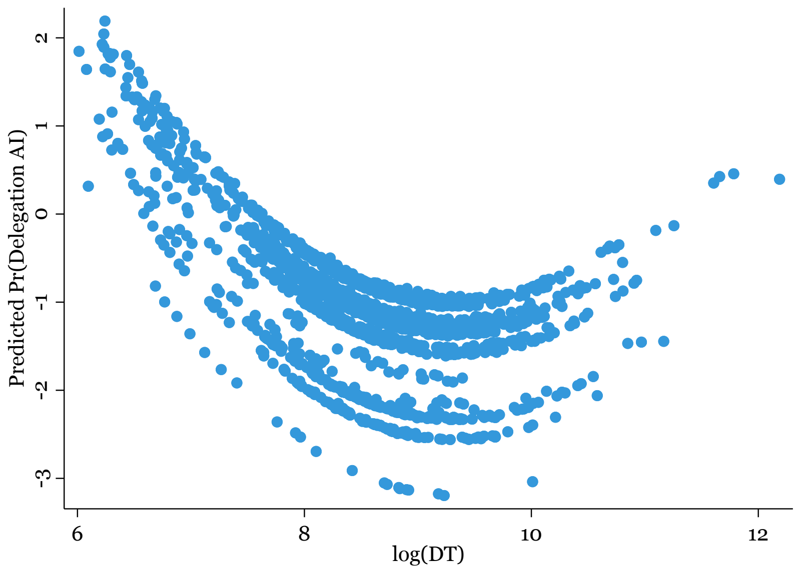
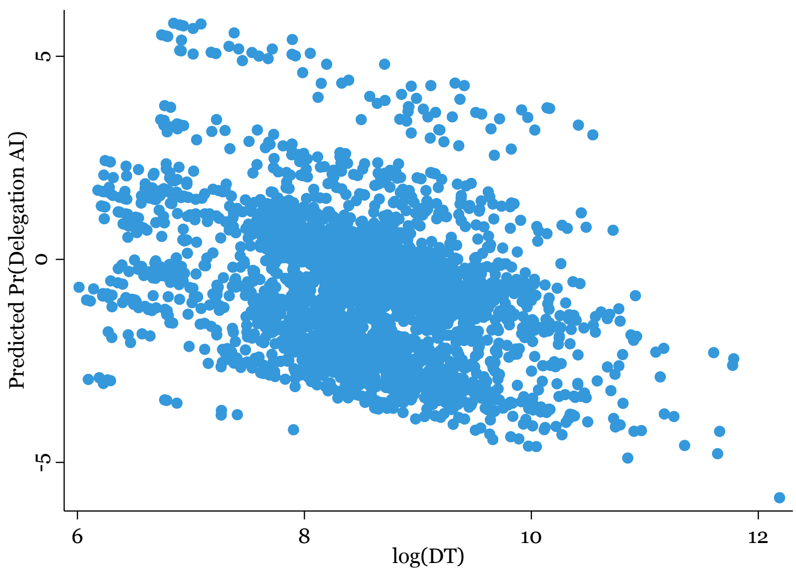


Figure 12.

The effect of log (DT) on predicted probability of delegating to AI for the coin context



References

Al-Natour, S., Benbasat, I., & Cenfetelli, R. (2011). The adoption of online shopping assistants: Perceived similarity as an antecedent to evaluative beliefs. *Journal of the Association for Information Systems*, *12*(5), 2.

Anderson, C. J. (2003). The psychology of doing nothing: forms of decision avoidance result from reason and emotion. *Psychological bulletin*, *129*(1), 139.

Arkoff, A. (1957). Resolution of approach-approach and avoidance-avoidance conflicts. *TheJournal of Abnormal and Social Psychology*, *55*(3), 402–404. https://doi.org/10.1037/h0043956

Ashoori, M., & Weisz, J. D. (2019). In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. *arXiv preprint arXiv:1912.02675*.

Bettman, J. R., Johnson, E. J., Luce, M. F., & Payne, J. W. (1993). Correlation, conflict, and choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 931.

Bland, M., Altman, D. G., & Rohlf, F. J. (2013). In defence of logarithmic transformations. *Statistics in Medicine*, *32*(21), 3766–3768. https://doi.org/10.1002/sim.5772

Bobadilla-Suarez, S., Sunstein, C. R., & Sharot, T. (2017). The intrinsic value of choice: The propensity to under-delegate in the face of potential gains and losses. *Journal of risk and uncertainty*, *54*(3), 187-202.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. Journal of Economic Behavior and Organization, 55, 467–484.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brockman, G., Murati, M., & Welinder, P. (2023). OpenAI API [Software]. https://OpenAI.com

Broniarczyk, S. M., & Griffin, J. G. (2014). Decision difficulty in the age of consumer empowerment. Journal of Consumer Psychology, 24(4), 608-625.

Butler, J. V., & Miller, J. B. (2018). Social risk and the dimensionality of intentions. *Management Science*, *64*(6), 2787-2796.

Caldwell, A. R., Lakens, D., Parlett-Pelleriti, C. M., Prochilo, G., & Aust, F. (2022). *Superpower* (Jade) [Software]. https://github.com/arcaldwell49/Superpower#the-anova_exact-function

Candrian, C., & Scherer, A. (2022). Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior*, *134*, 107308.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809-825.

Chatterjee, S., & Heath, T. B. (1996). Conflict and loss aversion in multiattribute choice: The effects of trade-off size and reference dependence on decision difficulty. *Organizational Behavior and Human Decision Processes*, *67*(2), 144-155.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668-1674.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of consumer research*, *24*(2), 215-231.

Dhar, R., & Sherman, S. J. (1996). The effect of common and unique features in consumer choice. *Journal of Consumer Research*, *23*(3), 193-203.

Dhar, R., & Simonson, I. (2003). The effect of forced choice on choice. *Journal of marketing research*, *40*(2), 146-160.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114.

Dyer, J. S. (2016). MAUT—multiattribute utility theory. In *Multiple criteria decision analysis: state of the art surveys* (2nd ed., Vol. 233, pp. 285–300). Springer.

Fasolo, B., Hertwig, R., Huber, M., & Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments?. *Psychology & Marketing*, *26*(3), 254-279.

Fischer, G. S., Carmon, Z., Ariely, D., & Zauberman, G. (1999). Goal-Based Construction of Preferences: Task Goals and the Prominence Effect. *Management Science*, *45*(8), 1057–1075. https://doi.org/10.1287/mnsc.45.8.1057

Frost, R. O., & Shows, D. L. (1993). The nature and measurement of compulsive indecisiveness. *Behaviour Research and Therapy*, *31*(7), 683-IN2. https://doi.org/10.1016/0005-7967(93)90121-a

Gul, F., Natenzon, P., & Pesendorfer, W. (2014). Random Choice as Behavioral Optimization. *Econometrica*, *82*(5), 1873–1912. https://doi.org/10.3982/ecta10621

Harris, J. G., & Davenport, T. H. (2005). Automated decision-making comes of age. *MIT Sloan Management Review*, *46*(4), 2-10.

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2020). *lab.js: A free, open, online study builder.* [Software]. https://lab.js.org/resources/publications/

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79*(6), 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

James H. Myers & Mark I. Alpert. (1977). Semantic Confusion in Attitude Research: Salience Vs. Importance Vs. Determinance. *ACR North American Advances*.

Kawaguchi, K. (2021). When will workers follow an algorithm? A field experiment with a retail business. *Management Science*, *67*(3), 1670-1695.

Kim, J., Novemsky, N., & Dhar, R. (2013). Adding small differences can increase similarity and choice. *Psychological science, 24*(2), 225-229.

Larrick, R. P. (1993). Motivational factors in decision theories: The role of self-protection. *Psychological Bulletin*, *113*(3), 440.

Leana, C. R. (1986). Predictors and consequences of delegation. *Academy of Management Journal*, *29*(4), 754-774.

Lloyd, J., Doll, H., Hawton, K., Dutton, W. H., Geddes, J. R., Goodwin, G. M., & Rogers, R. D. (2010). How Psychological Symptoms Relate to Different Motivations for Gambling: An Online Study of Internet Gamblers. *Biological Psychiatry*, *68*(8), 733–740. https://doi.org/10.1016/j.biopsych.2010.03.038

Logg, J. M. (2017). Theory of machine: When do people rely on algorithms?. *Harvard Business School working paper series# 17-086*.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Long, X., Sun, J., Dai, H., Zhang, D., Zhang, J., Chen, Y., ... & Zhao, B. (2022). The Choice Overload Effect in Online Recommender Systems: Theoretical Framework and Field Experiment. *Available at SSRN 3890056*.

Louviere, J. J., & Islam, T. (2008). A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research*, *61*(9), 903–911. https://doi.org/10.1016/j.jbusres.2006.11.010

Luce, M. F., Bettman, J. R., & Payne, J. W. (1997). Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 384.

Mahmud, H., Islam, A. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390.

Marcatto, F., & Ferrante, D. (2008). The Regret and Disappointment Scale: An instrument for assessing regret and disappointment in decision making. *Judgment and Decision Making*, *3*(1), 87–99. https://doi.org/10.1017/s193029750000019x

Neighbors, C., Lostutter, T. W., Cronce, J. M., & Larimer, M. E. (2002). Exploring college student gambling motivation. *Journal of gambling studies*, *18*(4), 361-370.

Owens, D., Grossman, Z., & Fackler, R. (2014). The control premium: A preference for payoff autonomy. *American Economic Journal: Microeconomics*, *6*(4), 138-61.

Oxford University Press. (2006). *Oxford Reference*. Oxford Reference. Retrieved March 31, 2023, from https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095426960;jsessionid=893CC6C3BA9F0E4C9C17B029A5585E1A

Persson, E., Erlandsson, A., Slovic, P., Västfjäll, D., & Tinghög, G. (2022). The prominence effect in health-care priority setting. *Judgment and Decision Making*, *12*(6), 1379–1391.

Rassin, E. (2007). A psychological theory of indecisiveness. *Netherlands Journal of Psychology*, *63*(1), 1–11. https://doi.org/10.1007/bf03061056

Rushworth, M. F. S., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*(6), 946–955. https://doi.org/10.1016/j.conb.2012.04.011

Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of consumer research*, *37*(3), 409-425.

Scholnick, Ellin K. and Clara S. Wing (1988), ''Knowing When You Don't Know: Developmental and Situational Consid- erations,'' *Developmental Psychology,* 24 (March), 190– 196.

Scholten, M., & Sherman, S. I. (2006). Tradeoffs and theory: The double-mediation model. *Journal of Experimental Psychology*. https://doi.org/10.1037/0096-3445.135.2.237

Schuett, J. (2019). A legal definition of AI. *arXiv preprint arXiv:1909.01095*.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: nonconsequential reasoning and choice. *Cognitive Psychology, 24,* 449-474.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*(1-2), 11-36.

StataCorp. (2016). *Stata: Release 16.* (16.1) [Software].

Steffel, M., & Williams, E. F. (2017). Delegating Decisions: Recruiting Others to Make Choices We Might Regret. *Journal of Consumer Research*, *44*(5), 1015–1032. https://doi.org/10.1093/jcr/ucx080

Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Tversky, A., & Shafir, E. (1992). Choice under Conflict: The Dynamics of Deferred Decision. *Psychological Science*, *3*(6), 358–361. https://doi.org/10.1111/j.1467-9280.1992.tb00047.x

Van Ittersum, K., Pennings, J. M., Wansink, B., & van Trijp, H. C. (2007). The validity of attribute-importance measurement: A review. *Journal of Business Research*, *60*(11), 1177–1190. https://doi.org/10.1016/j.jbusres.2007.04.001

Vass, C., Rigby, D., Tate, K., Stewart, A. K., & Payne, K. (2018). An Exploratory Application of Eye-Tracking Methods in a Discrete Choice Experiment. *Medical Decision Making*, *38*(6), 658–672. https://doi.org/10.1177/0272989x18782197
Wang, P. (2008, March). What Do You Mean by" AI"?. In *AGI* (Vol. 171, pp. 362-373).

Willemsen, M. C., & Keren, G. (2002). Negative-based prominence: the role of negative features in matching and choice. *Organizational Behavior and Human Decision Processes*, *88*(2), 643–666. https://doi.org/10.1016/s0749-5978(02)00006-7

Willemsen, M. C., & Keren, G. (2003). The meaning of indifference in choice behavior: Asymmetries in adjustments embodied in matching. *Organizational Behavior and Human Decision Processes*. https://doi.org/10.1016/s0749-5978(02)00521-6

Willemsen, M. C., Graus, M. P., & Knijnenburg, B. P. (2016). Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction*, *26*(4), 347-389.

Xu, J., Jiang, Z., & Dhar, R. (2013). Mental Representation and Perceived Similarity: How Abstract Mindset Aids Choice from Large Assortments. *Journal of Marketing Research*, *50*(4), 548–559. https://doi.org/10.1509/jmr.10.0390

Yukl, G., & Fu, P. P. (1999). Determinants of delegation and consultation by managers. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *20*(2), 219-232.