For What It's Worth

Humans Overwrite Their Economic Self-interest to Avoid Bargaining With AI Systems

Erlei, Alexander; Das, Richeek; Meub, Lukas; Anand, Avishek; Gadiraju, Ujwal

DOI
10.1145/3491102.3517734

Publication date
2022

Document Version
Final published version

Published in
CHI 2022 - Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems

For What It's Worth: Humans Overwrite Their Economic Self-interest to Avoid Bargaining With AI Systems

Alexander Erlei
alexander.erlei@wiwi.uni-goettingen.de
Georg-August-Universität Göttingen
Germany

Richeek Das
richeek@cse.iitb.ac.in
Indian Institute of Technology Bombay
India

Lukas Meub
lukas.meub@wiwi.uni-goettingen.de
Georg-August-Universität Göttingen
Germany

Avishek Anand
anand@l3s.de
Leibniz University Hannover
Germany

Ujwal Gadiraju
u.k.gadiraju@tudelft.nl
Delft University of Technology
Netherlands

ABSTRACT

As algorithms are increasingly augmenting and substituting human decision-making, understanding how the introduction of computational agents changes the fundamentals of human behavior becomes vital. This pertains to not only users, but also those parties who face the consequences of an algorithmic decision. In a controlled experiment with 480 participants, we exploit an extended version of two-player ultimatum bargaining where responders choose to bargain with either another human, another human with an AI decision aid or an autonomous AI-system acting on behalf of a passive human proposer. Our results show strong responder preferences against the algorithm, as most responders opt for a human opponent and demand higher compensation to reach a contract with autonomous agents. To map these preferences to economic expectations, we elicit incentivized subject beliefs about their opponent's behavior. The majority of responders maximize their expected value when this is line with approaching the human proposer. In contrast, responders predicting income maximization for the autonomous AI-system overwhelmingly override economic self-interest to avoid the algorithm.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; *User studies*; *Empirical studies in collaborative and social computing*; • **Applied computing** → *Economics*; Psychology.

KEYWORDS

AI system; Online Experiment; Human-AI Interaction; Decision Support System; Market Interaction; Ultimatum Bargaining

ACM Reference Format:
Alexander Erlei, Richeek Das, Lukas Meub, Avishek Anand, and Ujwal Gadiraju. 2022. For What It's Worth: Humans Overwrite Their Economic Self-interest to Avoid Bargaining With AI Systems. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3517734

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517734

1 INTRODUCTION

Many modern economic bargaining environments are experiencing a surge in decision-making based on artificial intelligence (AI). In consumer markets, algorithms determine the allocation of scarce goods through differential pricing [27, 29, 36, 37, 72, 73], automate financial transactions like credit-score loaning [107, 138, 151, 164] or trading [19, 32, 43, 74, 110], augment decision processes in credence goods markets [96], and are primed to autonomously act in a variety of online auction processes such as selling used goods [85, 130]. In organizations, AI-technologies can be leveraged to monitor, rate and reward employees [95], determine who gets hired [100, 109], and must function within interdisciplinary team environments that involve everyday negotiations such as allocating workloads [10, 70]. On a societal scale, AI might assist in driving tax policies [167] or judicial outcomes, thereby interfering in fundamental democratic negotiations between different civil factions.

To operate across such a vast, heterogeneous spectrum of bargaining scenarios, AI-systems need to be able to anticipate human behavior, learn from it, and react accordingly. This also entails an understanding of how the introduction of AI-systems will influence traditional decision patterns. Consequently, there has been a lot of behavioral research focusing on how humans make use of or rely algorithmic and AI-systems [42, 46, 114], or cooperate with them [41, 66, 117, 149, 168]. In many bargaining situations, however, people are not users or collaborators, but competitors, or simply the ones affected by the machine's decision. Rather than functioning as uni-dimensional decision aids, AI-systems are embedded in multidimensional human-human interactions where different stakeholders might assign them completely different roles.

For example, pricing algorithms essentially function as bargaining entities that largely substitute price-setting by the seller, and require the consumer to engage into economic negotiation with the system. When determining a price, the algorithm basically derives a take-it-or-leave-it offer about how to split the benefits of a transaction for a particular product, and the consumer either agrees to purchase the product, or rejects the bargain to go to a competitor. For example, consider an apple with production costs of $0.1, a

price of $1.5, and a consumer with a willingness to pay of $2. Here, the price functions as an initial negotiation offer whereby the seller secures a profit of $1.4 and the consumer retains additional $0.5 for other purchases. While primarily an economic conflict, it also entails a whole host of social considerations like distributive fairness or assertiveness, which exert different influence depending on the nature of the seller [83, 133, 137, 143]. Similar arguments can be made for automatic credit scoring, algorithmic hiring or evaluating and compensating employees. People without access to a system engage in economic bargaining with an algorithm, whereas a few years ago, they would have interacted with a human [129].

All these situations involve complex social dilemmas, which humans have been surprisingly good at managing. Rather than relying on purely pecuniary motives, humans have developed social skills and norms that consistently outperform game-theoretical expectations [56, 57, 80, 152]. We care about other people's payoff, broadly agree on context-dependent distributional norms like the 50/50-rule, reciprocate actions that signal trust, and feel obligated to fulfill responsibilities within teams [5, 6, 53].

To date, it is unclear if, and to what extent, these fundamentals of human behavior translate into human-algorithm interactions. Specifically, we have little data on how the introduction of algorithmic systems alters human beliefs, human needs and human behavior in general. These questions are of major importance, not only for the design of algorithmic systems, but more generally the desirability of and requirements towards their installation [135]. Going back to the pricing algorithm example, it has long been established that consumers' perceptions and acceptance of prices are partially dependent on distributional preferences towards the seller, i.e. the relation between consumer and producer surplus [89, 159]. We argue that these attitudes will likely change with the introduction of pricing algorithms. Indeed, first evidence of human-algorithmic interactions suggests that social concerns do change in comparison to purely human contexts, as people appear to be less emotionally involved [2, 121] and more demanding [52], both of which can hurt economic and social outcomes.

In light of the pervasiveness of bargaining algorithms in today's economies, this article aims to make a general contribution on the role of economic reasoning as a fundamental behavioral driver in human-algorithm interactions. By exploiting the controlled, abstract environment of an economic game, we aim to quantify human preferences for AI-systems in a decision context that entails economic stakes, and compare those preferences to an individual's economic self-interest. While AI-systems have the potential to increase market efficiency and economic welfare through improved predictions, these positive effects can only unfold in so far as the individual market actors (i) recognize that potential and (ii) subsequently engage with them. Disentangling economic motives from social motives derives causal evidence about the hurdles of AI-systems within bargaining environments and points towards potential structural changes in economic behavior. We therefore experimentally quantify human behavior when interacting with agents that embody varying degrees of AI-system agency to answer the following three research questions:

RQ1: Do humans self-select into specific interactions depending on an AI-system's autonomy?
RQ2: Are human preferences for or against AI-systems explained by economic expectations?
RQ3: Does the introduction of AI-systems into a bargaining environment affect overall economic welfare?

We exploit the controlled environment of an extended two-player ultimatum bargaining game. Ultimatum bargaining is commonly used to represent a wide range of competitive economic negotiating settings like price setting, employment negotiations or auctions. Two players bargain over a sum of money. Player one – the proposer – initially makes an offer on how to split that sum. Player two – the responder – either accepts or rejects that offer. If the offer is rejected, both receive nothing, if the responder accepts, the money is allocated according to the proposer's offer. The abstract nature of this task allows us to infer insights about a variety of mechanisms relevant to different societal actors. For example, **RQ1** might refer to a managerial dilemma where an organization decides on integrating algorithmic systems into their decision-making procedures. Employee preferences for or against AI-systems e.g. in evaluating job performance or hiring decisions can directly influence the system's effectiveness [20, 94, 102, 158]. Within a market environment, consumer preferences towards e.g. algorithmic pricing [22, 111, 159] relate to their purchasing choices, a firm's competitiveness, profits, overall economic efficiency, and consumer welfare. **RQ2** is of primary interest for theory building and predicting human behavior – a primary function of many real-world AI-systems –, while also allowing organizations to generate more precise cost benefit estimations. Finally, **RQ3** is relevant to regulators that aim for policies that increase overall economic welfare. If, for example, the introduction of AI-systems changes human reciprocal tendencies, a naive real-world implementation could decrease cooperation and hurt society at large. While the external validity of experimental research partially hinges on the assumption that human behavior does not systematically vary (too much) between the research environment and the outside world [108], there is a vast literature documenting the predictive power of abstract experimental findings for real-world decisions [14, 30, 35, 60, 71, 78, 92, 104, 156], although there can be experimentally unobserved confounds [62]. The ultimatum game specifically has been demonstrated to have high external validity outside the laboratory [68].

In our extended version of the ultimatum game, we first gauge human preferences for or against AI-systems with varying decision autonomy by letting responders choose to either bargain with a human, a human endowed with an AI decision aid, or an autonomous AI-system. Second, we map these preferences to economic expectations by eliciting incentivized responders beliefs about the behavior of each proposer type. In comparing responder beliefs, approach decisions and subsequent bargaining, we test whether human preferences for or against certain proposer types are rooted in economic concerns. This also allows us to directly infer whether behavioral patterns induced by the AI-systems are caused by social, human factors or merely reflect economic self-interest. Third, we elicit proposer beliefs towards responder behavior, analyzing whether

humans are able to anticipate potential behavioral changes induced by AI-systems. Finally, we compare aggregated responder and proposer behavior to draw inferences for economic welfare and market efficiency.

Our results show strong responder preferences against bargaining with an autonomous AI-system. This pattern cannot be explained by economic reasoning. Responders are more likely to expect the autonomous AI-system to maximize their income compared to both human proposer types. However, only a small minority of those subjects actually opt for interacting with the autonomous system. In contrast, the majority of responders who expect the human proposer to maximize their economic returns subsequently also choose to approach the human. Results for the human proposer supported by an AI decision aid fall squarely between the two poles. Furthermore, responders who do approach the autonomous AI-system ask for significantly higher compensation, reducing the share of successful interactions and thereby hurting overall economic welfare. These results provide strong evidence that humans are not only averse towards interacting with algorithmic systems, but that these aversions are grounded in fundamentally social factors which override economic self-interest. Thus, the introduction of algorithmic systems cannot simply be managed through monetary incentive schemes or the promise of efficiency gains. Rather, they introduce a fundamental shift in human decision-making, altering the motives for certain behavior and creating severe obstacles for real-life implementation.[1]

2 RELATED WORK

This article broadly relates to prior work about the effects of algorithms, AI-systems and machines on human behavior in social or strategic settings. Building on early experimental work within the *computers are social actors (CASA)* paradigm [123], a lot of research has shown that humans tend to treat machine actors as social agents by e.g. ascribing stereotypes [12, 87, 147], reacting to social cues [3, 44, 45], being triggered into social conformity [79, 139] or acting reciprocal [59, 93]. Yet, people also regularly demonstrate substantially less emotional involvement as well as lower social concerns when interacting with artificial agents. Both, behavioral data from economic games [1, 112, 121, 142, 148] and neurophysiological studies [34, 140, 154] point to fundamentally different social processes underlying human responses to artificial agents. Often, these tendencies exhibit substantial economic upside as people appear to act more rational and thereby increase cooperation [140] and market efficiency [119] or decrease the probability of herd behavior and bubbles [54]. Specifically in strategic settings, machine actors have been shown to increase cooperation in the prisoner's dilemma [84], group formation games [146] or bargaining [51, 86]. However, in domains where social norms regularly increase efficiency, machine actors often worsen outcomes. For example, humans are more likely to cheat when interacting with machines [39], cooperate less in public good games [161], and feel less remorse when exploiting machines [121]. Specifically in ultimatum bargaining, evidence suggests that computer players strongly reduce the impact of social concerns on behavior [13, 140, 154]. This effect appears to vanish once machines play on behalf of humans who still experience the economic consequences of the bargain. In such cases, people become more demanding of the machine, in particular when they experience feelings of unfairness [52]. Thus, overall, the experimental evidence points to a reduction in pro-social behavior as well as decreased importance of social norms in human-machine interactions. Still, a lot of ambiguity is left concerning the specific motives behind these changes, and human reactions are heavily context dependent. As illustrated for ultimatum bargaining, machines might provoke completely different reactions once they do not exist in an experimental vacuum, but take on the role of an economic agent that produces real consequences. Moreover, whether increases in economic rationality and self-interested behavior are caused by a lack of social motives, or different expectations regarding other players' decisions, remains completely open. Indeed, much experimental research uses computer players to *induce* rational expectations in human players. Therefore, behavioral changes might well be driven by changes in expectations, rather than the crowding-out of social concerns (see e.g. March (2021) [119] for an overview).

This article also contributes to the literature about what we call the *avoidance* of algorithmic and machine decision entities. There is growing evidence that in moral, social and subjective domains, humans judge machines to be largely unfit decision-makers. Bigman and Gray (2018) document that human aversions towards moral decisions by machines are consistent across a variety of social domains and in part caused by a perceived lack of mind [18]. Similar behavioral patterns have been observed for delegation in moral domains [63], subjective tasks [33], morally controversial investments [125], moral dilemmas and autonomous driving [48, 162], managerial [106] or medical decision-making [116]. Yet, when tasks are more objective [33, 106], humans lack expertise [114], humans retain some agency [47], humans and machines share some common rationale [144] or after prior exposure [99], people might also prefer or at least accept machine decision making. Particularly for strategic environments that combine mechanistic, rational game-theoretical expectations with social concerns, the literature thus does not provide any straightforward predictions.

We therefore provide three essential elements that add to the different strands of literature. One, we quantify aversions towards AI-systems with varying degrees of autonomy in a strategic environment that engenders both monetary self-interest as well as social concerns and beliefs. Second, we isolate the effect of economic self-interest by eliciting incentivized subject beliefs and thus provide counterfactual data on whether peoples preferences towards AI-Systems are economically or socially motivated. This also relates to the broader point about which fundamental drivers of behavior guide human decision-making in human-machine interactions. We do not measure specific social motives that might otherwise explain behavior. However, we do provide an exploratory qualitative analysis that reveals several potential non-pecuniary concerns to guide future research. Third, we measure whether those people who self-select into different bargaining situations subsequently also make different choices, and how these translate into overall welfare outcomes.

[1]We provide all data as well as screenshots from the experiment including the instructions via the following online repository: https://osf.io/7r486/?view_only=0e722d5e4e9748c2b30a380f63cc2659

3 EXPERIMENTAL DESIGN

Our experimental setup comprises three main decision stages. Subjects first indicate incentivized beliefs about their opponents behavior using the binarized scoring rule (BSR) [82]. Responders then decide which proposer to approach, and continue with the ultimatum bargaining. Proposers immediately proceed to the bargaining stage.

The experiment was conducted using the online platform *Prolific*[2] and accessible to native English speakers with a minimum approval rate of 90. We initially aimed for a total of 500 participants and finished with a sample of 480 (289 responders, 191 proposers) subjects due to attrition. We first gathered all responder observations, then the proposer observations, and later matched them randomly to determine payoffs.

Throughout the experiment, we used "Coins" as an experimental currency unit. Coins were later converted into Pound sterling, with a conversion rate of 1 Coin = 0.01£. Subjects earned a base payment of 1£ with a maximum bonus payment of additional 7.5£. The average proposer earned 4.9£ ($6.8) in about 17 minutes (17.3£/h), the average responder earned 4.8£ ($6.7) in about 20 minutes (14.4£/h). We provide screenshots of our experiment and the original instructions via the online repository (https://osf.io/7r486/?view_only=0e722d5e4e9748c2b30a380f63cc2659).

3.1 Ultimatum Game

The experimental design builds on the well-established ultimatum bargaining game from experimental economics [69], where it is widely used to analyze the tension between self-interested and pro-social decision-making. It is a core paradigm for a large and diverse field of behavioral social sciences. These include, among others, economics and game theory, psychology, evolutionary theory, biology or neuroscience [67, 152]. Its explanatory power for the role emotions, psychology or social concerns play in real-life negotiations has been validated across decades of research and numerous different cultures, occasionally highlighting how societal norms and identities outside WEIRD countries [77] change individual reactions toward unequal outcomes and social expectations regarding fairness and reciprocal behavior [64, 75, 76]. Yet, many core findings, such as the overwhelming rejection of highly unequal offers downward 20% of the pie, replicate almost anywhere, anytime. The clear differentiation between pure economic interest that can be mathematically derived and social concerns, in combination with its simplicity and transparency, make ultimatum bargaining an ideal framework to isolate the effect of economic rationale on social behavior. For example, prior studies from Neuroscience have examined human brain responses to unequal or unfair offers, establishing differential activation depending on the unfairness of an offer and, sometimes, the recipient. Bargaining with a computer or a random move elicits weaker, less emotional affective responses [50, 140, 155, 163]. Specific social motives that have been linked to decisions in the ultimatum game include reciprocity [90, 124] assertiveness [160], personality traits [166], Social Value Orientation [91], fairness concerns [127, 136], retaliation [25], or social comparison [21]. This heterogeneity within the vast literature on ultimatum bargaining reflects both the richness and complexity of a framework whose implications reach beyond negotiation into foundational matters of decision-making and rationality.

A proposer X offers the responder Y a certain fraction of a pie with size p. The proposer keeps x and the responder receives y, where $x, y \geq 0$ and $x + y = p$. The responder decides on a minimum offer z, where $z \geq 0$, and accepts the proposal by the proposer when the minimum offer is met by the proposer, i.e. if $y \geq z$ and $(x, y) = 1$. In case the minimum offer is not met by the proposer offer, i.e. if $z > y$, the responder rejects the offer and $(x, y) = 0$. Payoffs are determined by $\delta(x, y)x$ and $\delta(x, y)y$, i.e. if the responder Y rejects the offer, both earn nothing.

Solving the game solely based on pecuniary outcomes in a one-shot interaction without reputation implies that responder Y should accept all positive offers, which gives $\delta(x, y) = 1$ for $y > 0$. This follows the rationale that receiving something is better than receiving nothing.[3] Proposer X anticipates this, and consequently offers the minimal positive amount, leaving X with almost the whole pie p and Y little more than nothing.

In this experiment, we implement the strategy method variant of the ultimatum game. Responders independently indicate the minimum offer they are still willing to accept from the proposer. If the proposer's offer satisfies the responder's minimum offer, the money is split according to the proposer's offer. If the offer is lower, both players receive nothing. Following Trautmann and van de Kuilen (2015), instead of making an offer from a continuous action set, proposers and responders choose one of six potential Allocations (see Table 1) [150]. For example, if the proposer offered Allocation 2, and the responder chose Allocation 4 as a minimum offer, both would receive nothing. If the responder chose Allocation 2 as a minimum offer, and the proposer offered Allocation 4, the money would be split according to Allocation 4. Overall welfare increases with the share of successful interactions, irrespective of the specific distribution of economic gains. That is because each rejection leaves both players with an income of zero. Following the example from the introduction, selling the apple at a price of $2.50 as opposed to $1.50 would lead to a failed negotiation where the customer (willingness to pay: $2) does not buy the apple. Any price equal or below $2 would instead lead to a successful trade, thereby adding to overall economic welfare. At a price point of $2, the seller (i.e. proposer) exactly matches the minimum offer the customer (i.e. responder) is still willing to accept. This maximizes proposer welfare, but both players experience net-gains.

3.2 Approach Decision and Proposer Types

In our design, instead of being assigned to a specific opponent, responders choose autonomously which proposer type to approach. Hence, we observe self-selection behavior. This allows us to capture real-life analogous scenarios where competing entities implement solutions with varying degrees of AI autonomy and private actors like consumers or employees reveal their a priory preferences without preceding exposure. It thus also gives us a sense of potential market barriers to new AI technologies that are not already widely disseminated throughout industries.

[2] https://www.prolific.co

[3] While this is the weakly dominant strategy for Y, all distributions (x, y) can be established as equilibrium outcomes. For multiple equilibria consider a certain threshold $\tilde{y}$ for acceptance by the responder Y, such that $[(x, y), \delta(\tilde{x}, \tilde{y}) = 1]$ if $\tilde{y} \geq y$ and $\delta(\tilde{x}, \tilde{y}) = 0$ otherwise.

	Allocation 1	Allocation 2	Allocation 3	Allocation 4	Allocation 5	Allocation 6
Proposer	500	400	300	200	100	0
Responder	0	100	200	300	400	500

Table 1: Set of allocations in the ultimatum game.

Responders choose between (1) a human proposer, (2) a human proposer with a supporting AI-system or (3) a passive human substituted by an autonomous AI-system. If responders opt for (1), the ensuing bargaining replicates the well researched human-human interaction. However, being embedded in a market context where AI-systems replace human agents might alter standard beliefs and behaviors as observed in purely human interactions. For (2) and (3), responders and proposers receive the same description for the AI-system: *The AI-system (Machine-Learning) was trained using prior interactions of comparable bargaining situations and participant personality data.* Thus, subjects received some basic insights about the system, while not being overloaded with process-related information. This ensured a straightforward decision environment without confounds relating to the effects of technical details, degrees of interpretability or cognitive overload [81, 115]. In referring to participant personality data, we wanted to avoid perceptions of triviality or incompetency relating to the AI-system (see also section 3.4). While it is likely that human perceptions about the sophistication of an AI-system moderate behavior in economic decision contexts, such an analysis lies beyond the scope of this study. In real-world interactions, we argue that most people will ascribe some basic competency to AI-systems [e.g. 7, 88, 98, 118, 128, 145, 157, 165] – if only because there are almost no situations where it would be in a person's interest to install a malfunctioning algorithm.

Responders who were assigned role (2) and had the option to use the AI decision aid were allowed to inquire the system by submitting each of the six possible allocations. The system then provided proposers with two probabilities: (1) the likelihood that the allocation would be rejected and (2) the likelihood that the allocation was the expected-value-maximizing offer. We used a rule-based system based on the real probabilities from the responder sample to ensure the system's usefulness.[4]

When deployed as an autonomous AI-system in (3), the AI-system always chose the expected value maximizing offer (Allocation 3). Since responders did not receive feedback about the proposer's offer until the experiment was completed, this decision has no influence on responder behavior. It does maximize average overall income, which we judged to be the fairest outcome. Choosing the offer that maximizes expected proposer income is also the most realistic scenario, particularly within competitive market environments.

[4]Allocation 1: probability of rejection = 96.4%, probability of expected income maximization: 3.6%; Allocation 2: probability of rejection = 73%, probability of expected income maximization: 23.4%; Allocation 3: probability of rejection = 13.5%, probability of expected income maximization: 59.5%; Allocation 4: probability of rejection = 2.7%, probability of expected income maximization: 10.8%; Allocation 5: probability of rejection = 0.9%, probability of expected income maximization: 1.8%; Allocation 6: probability of rejection = 0%, probability of expected income maximization: 0.9%;

3.3 Belief Elicitation – Binarized Scoring Rule

One key aspect of our design is the differentiation between economic and social motives. To capture subjects' economic expectations regarding the three proposer types, we elicit their beliefs using the BSR. The BSR is a modification of the quadratic scoring rule that is independent of an agent's risk preferences. It is designed such that revealing once true beliefs about another player's behavior maximizes the probability to earn a fixed reward. In our implementation, the fixed reward amounts to 250 Coins. Subjects indicate their beliefs about their opponents behavior for a pre-determined set of conditions. Responders always indicate their belief that the offer from a given proposer satisfies a given Allocation. Proposers indicate their belief that the responder's minimum offer satisfies a given Allocation.

Row No.	Probability of reward if Responder accepts Allocation #	Probability of reward if Responder rejects Allocation #
1	1.0000	0.0000
2	0.9975	0.0975
3	0.9900	0.1900
4	0.9775	0.2775
5	0.9600	0.3600
6	0.9375	0.4375
7	0.9100	0.5100
8	0.8775	0.5775
9	0.8400	0.6400
10	0.7975	0.6975
11	0.7500	0.7500
12	0.6975	0.7975
13	0.6400	0.8400
14	0.5775	0.8775
15	0.5100	0.9100
16	0.4375	0.9375
17	0.3600	0.9600
18	0.2775	0.9775
19	0.1900	0.9900
20	0.0975	0.9975
21	0.0000	1.0000

Table 2: Proposer Decision table

Subjects receive the fixed reward if their realized score is smaller than a number drawn from a uniform distribution [0,1]. To determine the score, we implement the following loss functions for responders: $(1-p)^2$ if $y \leq z$ and p^2 if $y > z$, with y being the responder payoff specified by the respective allocation, z being the income according to the proposer offer, and p giving the probability that the proposer offer satisfies the responder demand. For proposers, we use $(1-p)^2$ if $z \geq x$ and p^2 if $z < x$ as the loss function.

Recall that responders receive the fixed reward when the randomly drawn number is greater than the score equivalent to the loss calculated. Thus, the reward is received with a probability $1-(1-p)^2$ if $y \leq z$ and $1-p^2$ if $y > z$.

Responders make eighteen decisions in total – one for each possible allocation and proposer type. This later allows us to compare their expectations regarding the different proposer types. Responders learn that at the end of the experiment, the computer randomly draws one of the eighteen scenarios to determine their payoff. Thus,

they know that each decision might be the one determining whether they receive the bonus of 250 Coins.

Instead of directly reporting their subjective probabilities, subjects choose one of 21 rows in a scoring table that calculates the respective probabilities for values of $p \in [0, 0.05, ..., 0.95, 1]$ (see example Table 2). Consider a clairvoyant responder who is asked to state a probability that a certain proposer's offer meets a minimum offer of 100 Coins (Allocation 2). In order to maximize the probability to receive the fixed reward, the clairvoyant responder should choose row one because it guarantees the reward. Choosing e.g. row six only yields a probability of 93.75%. However, if the responder weren't clairvoyant, row six would also set a probability of 43.75% if the proposer's offer turned out to be lower than the minimum offer of 100 Coins (Allocation 2), which is a lot better than the probability of 0 as determined by the first row in case of rejection. Accordingly, choosing a row in the middle indicates uncertainty about whether a specific proposer type will meet the respective allocation as it gives a fair chance to earn the fixed bonus for both possibilities. Choosing a high row number indicates relative certainty that a specific proposer type will not meet the respective allocation. Overall, consistent beliefs should move from lower rows for allocations indicating a greater share of the pie for proposers to higher rows for allocations indicating a greater share of the pie for responders. Conditional on the proposer type for a specific allocation, higher row numbers indicate that responders believe this proposer to be less likely to met the respective minimum offer and thus being less generous.

Proposers also indicate their beliefs for the eighteen scenarios – one for each possible combination of allocation and proposer type. They only learn their specific proposer type after the belief elicitation stage. Thus, at the belief elicitation stage, each of the eighteen decisions might be the one determining the probability of receiving the fixed reward. Equivalent to responders, proposers state for each scenario their subjective likelihood that the responder's minimum offer at least equals the scenario's proposer offer. Thus, proposer beliefs should tend to diametrically oppose responder beliefs.

3.4 Procedure

Upon voluntary enrollment, responders first completed a simple attention check and completed the Social Value Orientation (SVO) Questionnaire [122], the negative reciprocity sub-scale from Perugini et al. [132] as well as a battery of demographic questions. We gathered personal data at the beginning of the experiment to allow for more sophisticated AI-system perceptions. Without any additional information, subjects might have thought the system to be trivial and therefore exhibit behavioral patterns that lie beyond our main analysis. Since both the SVO orientation and the negative reciprocity scale have been empirically connected to ultimatum bargaining [17, 24, 91], we expect participants to recognize them as meaningful. This was confirmed by the post-experimental questionnaire, where over 85% of subjects agreed (6% disagreed) that *"the information stated at the beginning of this task say something about me as a person"* on a seven point Likert-scale.

Responders then proceeded to the ultimatum game instructions, after which they had two trials to answer three comprehension questions correctly. Those who failed both trials were prevented from proceeding with the study. Next, responders read through the belief elicitation instructions. Again, they had to answer two comprehension questions within two trials in order to proceed to the main experiment. We opted for this strict selection mechanism because mathematical belief elicitation schemes necessitate a careful reading of the instructions due to their inherent complexity. This is particularly important for online environments, since those can suffer from a lack of subject attention [61].

In the main experiment, responders first learned their role and subsequently indicated their beliefs for all 18 scenarios using the scoring table. After a short reminder screen, responders proceeded to the approach decision. After choosing between the three proposer types, they indicated the minimum allocation they were still willing to accept from the priorly selected proposer, and finally completed the post-experimental questionnaire. Subjects were only informed about the proposer offer after the experiment was completed, and did not receive feedback during the belief elicitation stage.

Proposers completed a similar, albeit shorter experimental procedure. After enrollment, they immediately proceeded with the instructions for the ultimatum game and the belief elicitation. Like responders, they had to pass the two blocks of comprehension questions. Following this, they learned about their proposer role, but not the specific proposer type. We did that to ensure that every one of the eighteen belief scenarios could be the one determining their final bonus and therefore guarantee incentive-compatibility.

Proposers then indicated eighteen beliefs, learned about their type, and chose their offer. Finally, proposers completed a short post-experimental questionnaire including a battery of demographic questions. Like responders, proposers did not receive information about their responder's minimum offer until the experiment was completed.

3.5 Participants

Table 3 summarizes participant demographics for responders and proposers. Our sample is predominantly female, relatively young, and mostly college-educated. Most subjects are either working, or looking for work, with very few postgraduate degrees.

	Responder	Proposer	Proposer + AI
female	78%	53%	67%
mean age	24.9	26.2	26.9
paid employee	52%	58%	61%
self-employed	7%	6%	6%
looking for work	19%	15%	12%
not working	18%	15%	18%
bachelor's degree	36%	34%	29%
college no degree	29%	14%	28%
high school grad	20%	28%	16%
master's degree	9%	15%	15%
N	289	97	94

Table 3: Participant demographics. For clarity, we omit categories with less than 5% frequency. A complete list is available in the online repository.

4 RESULTS

We first describe the general results of the bargaining stage, and then dive deeper into the analysis of responder behavior. We isolate the effect of economic beliefs on responder choices for the three proposer types. Further, we check whether proposers anticipate heterogeneous responder behavior conditional on the proposer type.

4.1 Approach and Bargaining

Responders exhibit a clear pattern of avoiding the autonomous AI-system (see Table 4).

Only a small minority selected the autonomous system playing on behalf of a human proposer, and the distribution of approach decisions is significantly different from a hypothetical uniform distribution ($\tilde{\chi}^2(2) = 61.7, p < .001$). Responders approached the human more than the human with the AI decision aid (binomial probability test: $p < .001$), and the human with the AI decision aid more than the autonomous AI-system (binomial probability test: $p < .001$). Still, frequencies are relatively similar between the two human proposers, while collapsing for the autonomous system. Our results show that responders forgo the autonomous AI-system in favor of both human alternatives. Preferences towards the AI decision aid are comparatively inconclusive.

> **Result 1:** Humans avoid interacting with autonomous AI-systems and prefer human opponents.

Furthermore, as shown in Figure 1, responders asked for significantly different minimum offers depending on which proposer type they approached (H(2) = 8.46, p = .014). In particular, subjects who bargained with the autonomous AI-system demanded significantly more than subjects who bargained with a human (H(1) = 6.87, p = .009) or the human with an AI-system (H(1) = 7.53, p = .006). As illustrated in Figure 1, this was predominantly because responders often demanded more than 50% from the autonomous AI-system (39%), which seldom happened for the human proposer (15% and 13% respectively) and points to a basic shift in either economic expectations or social concerns.

> **Result 2:** Humans who self-select into bargaining with an autonomous AI-systems subsequently ask for higher compensation.

From 191 active proposers, 97 were assigned the role of human proposer, and 94 were endowed with the AI-system decision aid. There are no differences in proposer offers (239 vs. 227; H(2) = 2.56, p = 0.109). Both types offered significantly more than the 200 Coins offered by the autonomous system (human: $t = 6.57, p < .001$; human + AI-system: $t = 4.51, p < .001$).

> **Result 3:** Proposer offers do not significantly change with the AI decision aid.

Comparing proposer and responder behavior conditional on the proposer type allows us to make some general inferences about market efficiency. Responders who engage with the autonomous system hurt ceteris paribus overall cooperation and efficiency due to more assertive, economically demanding decisions that decrease the probability of a successful interaction. While responders acted more assertively towards the autonomous AI-system, the system itself was less generous than the average human proposer. Since it was primarily informed by human-human interactions, always offering 200 Coins and thus a 40/60 share would have increased aggregated income compared to the human proposer types. However, as responders changed their behavior and became more assertive, the system was actually less effective than the human proposers. In fact, the maximum market efficiency for human proposers and human proposers with the AI decision aid is almost 100%. If matched correctly, all responder demands except one could have been satisfied by all proposer offers (see Figure 1). In contrast, 39% of responder minimum offers would have never been matched by the autonomous AI-System. Among others, this highlights the need for AI-systems to learn from real-world data that incorporates real interactions specifically between the system and its human environment. Learning from human-human interaction proves inefficient and leads to sub-optimal outcomes.

> **Result 4:** The introduction of an autonomous AI-system reduces total market potential due to more assertive responder behavior.

Using multinomial and ordered logistic regressions, we find that both responder approach behavior and minimum offers are largely independent from demographic data.[5] Subjects with self-reported doctoral or professional degrees (JD, MD) tended to demand more money, while Bachelor, Master and high school degrees were associated with lower offers. Subjects with SVO type competitiveness stated significantly higher minimum offers (OR = 74.84, p = .02), but only two individuals fell into that category. For proposer offers, there are no significant or consistent trends. Pairwise chi-square comparisons of responder approach decisions and subjects employment status ($\tilde{\chi}^2$ (12) = 8.07, p = .78), sex ($\tilde{\chi}^2$ (2) = 0.38, p = .83) or highest degree received ($\tilde{\chi}^2$ (14) = 22.56, p = .07), reveal no significant relationship. The data on education attainment potentially suggests that preferences for humans might be exacerbated in college-educated individuals, as high school graduates (N = 58) were the only group with aggregate preferences for the AI support system (55%) over the purely human proposer (38%). Overall, our results are robust towards demographic data.

4.2 Responder and Proposer Beliefs

One main goal of this article is to analyze whether economic expectations can explain differences between human-AI and human-human interactions. This section looks at belief data to parse out the economic from the social motivations. After some aggregated statistics, we primarily analyze responder within-subject behavior

[5] For the regression tables on demographic data, please refer to the online repository.

		Human Proposer	Human Proposer + AI-system	Autonomous AI-system	Total
approaches	#	142	111	36	289
	%	49.1	38.4	12.5	100
minimum offer	mean	190	186	228	193
	sd	65.6	79.2	84.9	74.5

Table 4: Summary statistics for responder behavior

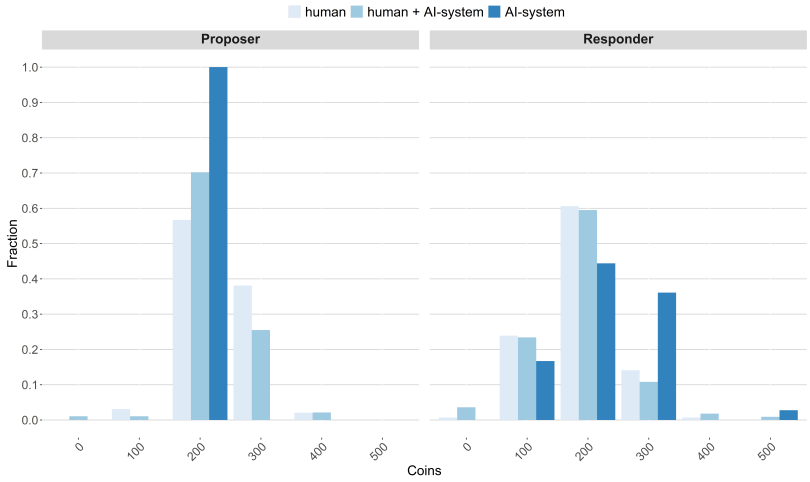

Figure 1: The fraction of proposer offers and responder minimum offers conditional on (the approached) proposer type.

to measure whether the introduction of AI agents fundamentally changes the explanatory patterns behind human behavior.

4.2.1 Responder Beliefs. Looking at aggregate belief patterns, we first test for basic coherence in responder data. We categorize a subject's beliefs as *unreliable* when for any of the three proposer types, they indicated a higher row number for Allocation 1 scenarios than Allocation 6 scenarios. Remember that indicating a higher row number for Allocation 1 than for Allocation 6 essentially manifests the belief that it is more likely for a proposer to satisfy a minimum offer of 500 than it is for them to satisfy a minimum offer of either 0, 100, 200, 300, 400 or 500. Thus, it is likely that these subjects either did not properly understand the task, or lacked some care in their answers for at least one proposer type. We find that 103 (35.6%) responders fall into the *unreliable* category, with many of them apparently confusing the two extremes as indicated by small local maxima at the highest row number where responders should have indicated the lowest number. While this share is high, it is not entirely unexpected, since these tasks are very complicated and we know of no other study that tried to implement them in an online environment as of yet. Throughout this section, we therefore always differentiate between the whole sample and the *reliable* sample (N = 186) when analyzing belief data, to ensure that our behavioral interpretations are robust to (1) subject understanding of the task and (2) possible selection bias due to factors correlating with an improper understanding of the task. As we will show, all main results in this paper are consistent with both samples.

Table 5 lists reported responder beliefs about whether the proposer's offer would satisfy the minimum offer specified in the respective allocation scenario. Aggregated results are consistent across proposer types and in line with prior studies on belief elicitation in ultimatum bargaining [150].

Allocation	Probability Human Proposer	Probability Human Proposer + AI-system	Probability Autonomous AI-system
(0, 500)	92% (17)	89% (19)	86% (24)
(100, 400)	77% (18)	76% (18)	74% (22)
(200, 300)	59% (20)	58% (19)	60% (22)
(300, 200)	40% (22)	40% (22)	42% (21)
(400, 100)	22% (22)	22% (20)	27% (22)
(500, 0)	9% (22)	11% (21)	14% (24)

Table 5: Summary statistics average responder beliefs – *reliable* sample. Standard deviations in parentheses. Allocation: (y, x)

This confirms that on population level, responders did not predict more economic exploitation from the autonomous AI-system. In contrast, they were even somewhat more likely to expect relatively high returns above the equal split. However, aggregate statistics conceal heterogeneous behavioral patterns on the individual level. Thus, we now turn to the within-subject analysis.

> **Result 5:** On the population level, there are no substantial differences in economic a priori beliefs about the three proposer types.

An easy way to measure the importance of economic rationale for responder bargaining choices is to determine whether responders

actually approach the proposer type that maximizes their expected income.[6] Therefore, we first determine a subject's expected-value maximizing choice by multiplying their self-reported beliefs with the respective income specified by each scenario. For example, a responder who believed the human proposer to match Allocation 2 (200 Coins for the responder) with a probability of 84% (row number 9) and the autonomous AI-system to accept the Allocation with a probability of 75% (row 11) has an expected income of 168 Coins for the human proposer and 150 Coins for the autonomous system.[7] Table 6 shows the same basic pattern across four samples. Contrary

	Human Proposer	Human Proposer + AI-system	Autonomous AI-system
max $\mathbb{E}(y)$: Whole Sample	130	129	155
max $\mathbb{E}(y)$: *Reliable* Sample	77	79	106
max $\mathbb{E}(y)$: One Dominant Choice	62	54	85
max $\mathbb{E}(y)$: *Reliable* One Dominant Choice	35	34	66

Table 6: Number of responders for whom each respective proposer type maximized their expected income. "One Dominant Choice" refers to the sub-sample where one proposer type dominated the other two and there are no ties.

to the behavioral data, responders are consistently more likely to expect the autonomous agent to maximize their economic returns compared to the other two alternatives.

Figure 2 depicts the share of responders who decided to approach the expected value maximizing proposer type. Both for the whole and the *reliable* sample, responders exhibited substantially different behavioral patterns depending on which proposer maximized their expected value.[8]

Overall, 52% (56%) approached the human proposer when they maximized the responder's expected income, 34% (33%) approached the human proposer with the AI-system, and only 12% (12%) approached the autonomous AI-system (*reliable* in parentheses). To make statistical inferences, we concentrate on the sub-sample of subjects who believe that one proposer type maximizes $\mathbb{E}(y)$ (*One Dominant Choice*; N = 201). While we lose some statistical power, it is necessary to avoid multiple occurrences of one individual and therefore guarantee non-overlapping, independent observations. First, nothing changes in terms of likelihood to approach the expected value maximizing proposer type: 47% (57%) human, 28% (26%) human with an AI-system, 14% (12%) autonomous AI-system. The difference is significant both for the whole ($\tilde{\chi}^2 = 19.02, p < .000$) and the *reliable* ($\tilde{\chi}^2 = 23.32, p < .000$) sample. There are no demographic differences.

Figures 3 provides an illustration of the coherence between a responder's EV-maximizing proposer type, and their subsequent approach decision for the *reliable* One Dominant Choice sample. Only a minority of subjects who approached the human proposer thought that they would maximize their economic returns. Those who believed the autonomous AI-system to maximize expected income spread relatively evenly across the two human alternatives. Many responders economically favoring the human with an AI decision aid also switched to the sole human proposer. This suggests that without an autonomous AI-system, preferences for the human over the AI decision aid might be more pronounced.

> **Result 6:** The introduction of AI-systems changes human behavior. People are less likely to follow economic rationale both when an AI decision aid or an autonomous agent maximizes their expected income. The effect increases with the AI-system's decision autonomy.

We show that responders appear to follow fundamentally different behavioral drivers depending on the bargaining opponent. When responders predict the human to maximize their income, the majority approaches them. When responders predict the autonomous AI-system to maximize their income, only a small minority chooses to bargain with it. Hence, while economic expectations have a lot of predictive power for human-human interactions, their importance collapses in human-agent bargaining. Instead, people appear to overwrite their economic self-interest in order to avoid the AI-system. Looking at the difference between humans, humans with an AI decision aid and humans being replaced by an autonomous system, this process appears to be moderated by the degree of autonomy the system inhibits. Endowing another human with an AI decision aid suffices to reduce the importance of economic rationale for responder decision-making, but the effect is significantly smaller than when the responder faces an autonomous AI-system.

Our results also confirm that – by elimination – human aversions towards algorithmic decision systems are caused by the social, human factors of a decision environment. To capture subjects' perception of unfairness towards the two AI-systems, we asked them to state their agreement with two statements on a seven point Likert-scale: (1) *I think it is unfair that one proposer gets to use a decision-support system* and (2) *I think it is unfair that one proposer has an autonomous agent playing on their behalf.* For the autonomous agent, 60% indicated perceptions of unfairness. However, this did not translate into different choices, as both groups were similarly unlikely to approach the system when thinking that it maximizes their expected income (unfair: 13%, not unfair: 11%). For the AI decision aid, 38% judged it as unfair, and differences between the two groups were more substantial (unfair: 27%, not unfair: 38%), but still insignificant ($\tilde{\chi}^2 = 1.66, p = .197$). Results do not change for the *reliable* sample. However, looking only at responders with one clear expected value maximizing proposer type suggests that, absent ambivalence, people appear to be more likely to follow economic self-interest when they do not judge the AI decision aid as unfair (unfair: 13%, not unfair: 39%; $\tilde{\chi}^2 = 4.34, p = .037$). For the autonomous AI-system, there is no such correlation ($\tilde{\chi}^2 = 0.73, p = .391$).

> **Result 7:** Overall, fairness perceptions largely appear to be unable to explain our results, as the effects towards a human with a decision aid are marginal, and non-existent for the autonomous AI-system.

[6] Note that we do not make an argument about the *rationality* of responder decisions. Economic self-interest does not necessarily determine rationality, as people e.g. might derive additional utility from interacting with other humans, or project future disutility from actions that promote AI-systems over humans in the long run.

[7] Results hold when using the average expected value over all six scenarios per proposer type, rather than the one proposer that provides the highest expected value.

[8] This pattern holds for all four sub-samples.

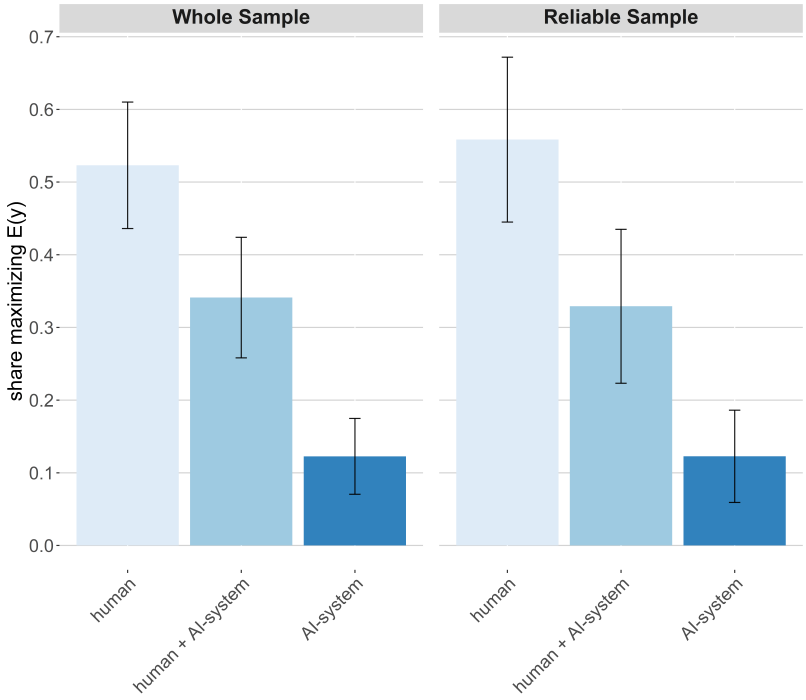

Figure 2: Share of responders approaching the EV-maximizing proposer type

4.2.2 Proposer beliefs. Table 7 shows proposer beliefs from the BSR for each allocation and proposer type regarding the probability that a responder accepts a given split. Like before, we create a *reliable* sample (N = 105) without subjects whose answers for at least one proposer type were unreliable. For proposers, answers were classified as unreliable when the indicated row number for Allocation 6 was higher than the row number for Allocation 1, following the same intuition as above. Proposers were substantially more likely to believe that their opponent would accept the largest unequal split (i.e. Allocation 3) than responders, which is in line with the experimental literature [67] and evidence for people's nuanced understanding of social situations.

Allocation	Probability Human Proposer	Probability Human Proposer + AI-system	Probability Autonomous AI-system
(0, 500)	8% (21)	11% (22)	11% (22)
(100, 400)	28% (19)	28% (19)	28% (23)
(200, 300)	60% (19)	55% (18)	52% (22)
(300, 200)	74% (19)	72% (18)	68% (20)
(400, 100)	86% (17)	85% (16)	85% (16)
(500, 0)	91% (22)	93% (16)	92% (17)

Table 7: Summary statistics average proposer beliefs – *reliable* sample. Standard deviations in parentheses. Allocation: (y, x)

Proposers also correctly anticipated that responders would be more demanding when bargaining with an autonomous AI-system as opposed to a human. While there are no overall differences for Allocation 1 ($t = 0.21, p = .831$), 2 ($t = 0.75, p = .456$), 5 ($t = 0.95, p < .001$) and 6 ($t = 1.17, p = .243$), beliefs percentages for Allocation 3 ($t = 4.49, p < .001$) and 4 ($t = 3.77, p < .001$) are significantly smaller. As shown above, this region around the equal split switching point is precisely where responders become more demanding towards the autonomous AI-system.

For the AI decision aid and the autonomous system, the difference is not significant for Allocation 3 ($t = 1.77, p = .078$) but for Allocation 4 ($t = 2.74, p = .007$). There are no other discrepancies. Nothing changes in the *reliable* sample.[9]

In line with these results, 82% of proposers stated after the experiment that they believed the responder would take the AI-system into consideration when deciding which proposer to approach. Resembling actual responder behavior, 45% of proposers expected responders to approach the human, 42% the human with the AI-system decision aid, and only 13% the autonomous AI-system.

> **Result 8:** Proposers correctly anticipate that responders become more economically demanding when bargaining with an autonomous AI-system.

When asked, if given the choice, which proposer type they would have selected, the majority of proposers (68%) favored the AI-system decision aid, 23% wanted to bargain on their own, and only 10% would have chosen the autonomous system. Thus, proposers did not only expect responder aversions towards the autonomous AI-system, but were broadly in agreement with their opponents. Our

[9]Human vs. AA: Allocation 1 ($t = 1.85$, $p = .067$), Allocation 2 ($t = 0.09$, $p = .924$), Allocation 3 ($t = 3.66$, $p < .001$), Allocation 4 ($t = 2.87$, $p = .005$), Allocation 5 ($t = 0.88$, $p = .383$), Allocation 6 ($t = 0.46$, $p = .646$); AI-system Decision Aid vs. AA: Allocation 3 ($t = 1.40$, $p = .164$), Allocation 4 ($t = 2.27$, $p = .025$)

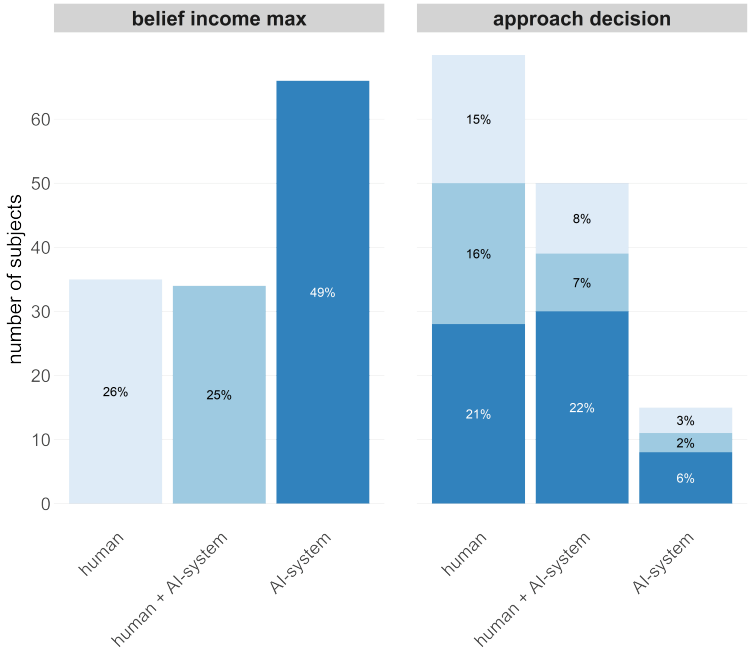

Figure 3: Left: The share of participants who expected a certain proposer type to maximize their income. Right: The composition of the approach decisions conditional on which proposer type subjects expected to maximize their income.

results shed light on some basic challenges for the real-life implementation of autonomous AI-systems within competitive markets or organizational environments. People who face an AI-system fundamentally change their behavior to avoid interacting with it, while those tasked with substituting another human actor anticipate these reactions – at least partially.

4.3 Exploratory Qualitative Analysis

At the end of the experiment, we asked responders: *"please explain why you chose to approach [proposer type]"*. We opted for an inductive manual coding scheme to identify responder response labels and gain insights into their reasoning [113]. Results from two independent coders were compared for consistency by the authors and amended if needed. There are two important caveats to the analysis. First, it relies on non-incentivized self-reported data near the end of the experiment. Thus, we cannot verify that subjects reflected carefully on their answers, and some could have used the question to justify their choices ex post. Second, we do not specifically ask for a responder's goal, but rather the mechanisms behind their decisions. As reflected in the actual data, this means that we can approximate different subject motivations, but do not gain consistent insights into responders' objective functions. We therefore propose the following results as a first step towards understanding why people might react differently to AI-systems, without claiming that they have any bearing on their actual utility functions.

In total, we identified 311 reasons from 289 responders. These 311 reasons were categorized into five distinct categories: empathy, fairness, predictability, information set and other. Answers categorized into *empathy* reflected that the responder values a proposer who includes emotions or empathy when deriving an offer, for example: *"I think this is the option that would allow for human emotion to be included"*. Answers in *fairness* emphasized preferences for either a fair, or an equal allocation. Examples are *"I thought the AI system might make their decision more fair; I do not necessarily trust a person working alone to understand the task or make a generous decision"* or *"I feel like the human could be more equal in decision making"*. *Predictability* captured responders who stated preferences to bargain with the proposer type whose decisions they could best predict, e.g. *"I feel as a human I can better estimate the actions of another human as opposed to AI"*. Finally, answers labelled as *information set* referred to responders valuing proposers who could draw from either the largest amount of information, or a specific kind of information. Examples include *"I feel that the emotional characteristics of a human being, combined with the logic and learned human behaviours of an AI, would give me the best possible outcome based on my minimum offer"* or *"I approached the autonomous agent because its decision making will be based off of other responders minimum splits, and might make a decision less informed by greed and more informed by logic"*. Answers that did not fit either one of the above categories were labelled as *other* (10%).

Figure 4 illustrates subject answers conditional on their approached proposer type. Roughly one-third of responders approaching the human emphasized an aspect relating to the proposer's empathy or emotions. In contrast, virtually no one indicated similar

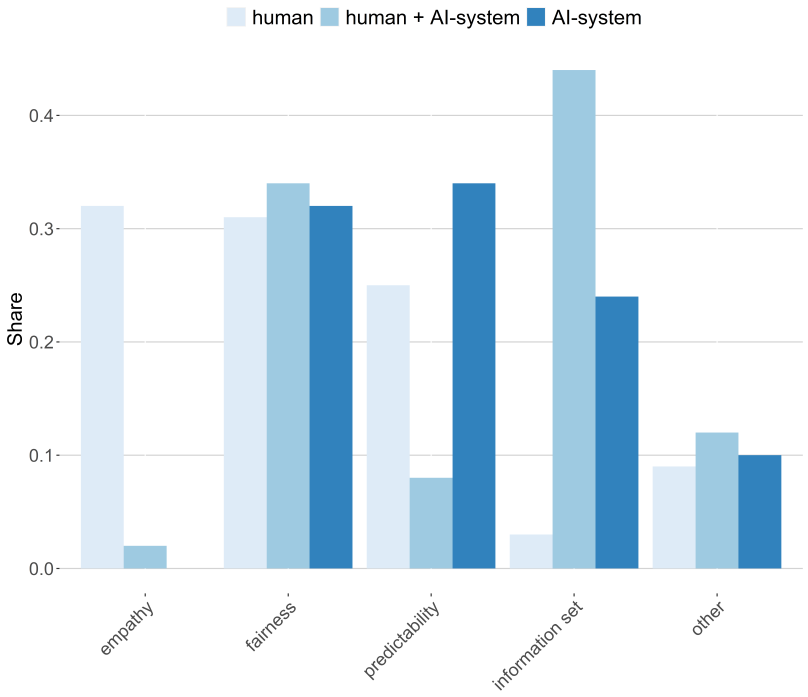

Figure 4: Categorization of responder answers to the post-experimental question *"please explain why you chose to approach [proposer type]".*

motivations when approaching the human with the AI support system or the AI-system alone. Instead, responders approaching the hybrid proposer type valued the shared information set, ostensibly combining distinct AI and human knowledge. The *predictability* category posits one potential downside of these perceptions. Compared to both the sole human and autonomous AI-system, the support system almost never elicited any reference towards increased predictability. This, instead, is where the AI-system strongly overperforms. In line with our analysis above, there were no differences in fairness statements.

These results reveal some insights into divergent responder decision processes. People appear to assign heterogeneous, but distinct judgments to each proposer type. Depending on the interplay of these perceptions and a responder's preferences, different behavioral patterns might arise. Here, it seems that subjects who value a maximization of available information tend to opt for the human who is assisted by an AI-system, whereas those who value shared humanness through empathy or emotions favor the sole human. Contrary to the other two, the autonomous AI-system does not have a clear comparative advantage in any category, although predictability and available information appear to be important determinants of approach decisions.

5 GENERAL DISCUSSION AND FINDINGS

The increasing dissemination of algorithmic decision-making systems leveraging AI disrupts and augments traditionally human environments, necessitating a careful analysis of potential behavioral changes. We provide first evidence that people overwrite their economic self-interest to interact with another human over an AI-system. While the effect is already measurable for decision aids, it is by far most pronounced for autonomous AI-systems that decide on behalf of another human. Our results therefore suggest that the introduction of AI-systems causes a general shift in the determinants of human behavior, where social concerns lead to systemically avoidant behavior towards AI-systems.

RQ1: Do humans self-select into specific interactions depending on an AI-systems autonomy? Participants in ultimatum bargaining overwhelmingly prefer to approach human proposers. While the human proposer without any assistance is the most prominent option, many responders also opt for the human with an AI decision aid. In contrast, the autonomous AI-system is strikingly unpopular. Responders self-selecting into bargaining with a human do not behave differently when the proposer has the option to use an AI decision aid. However, those who choose the autonomous AI-system ask for significantly more compensation. This effect appears to be primarily driven by a large increase of demands that exceed an even split. These results cannot be explained by stated fairness perceptions. While many responders judge both the decision aid and the autonomous AI-system as unfair, their behavior differs only marginally compared to those who did not perceive the systems as unfair. It is possible that both the selection effect, and increased responder demands are – at least partially – elicited by the apparent asymmetry between the two roles. Responders without access to a

system might act more assertive to compensate for a perceived lack of agency or power. However, it seems unlikely that people who feel disempowered by the AI-system would choose to approach it. Thus, anticipating a sense of impotence might be more useful in explaining selection behavior. Some prior evidence suggests that responders adjust their demands upwards when they know that proposers are ex ante informed about their specific minimum offer [134]. Similarly, responders who bargain with an AI-system might act more in accordance with what they expect the system to predict about them. Self-serving bias could lead responders to attribute assertiveness to their own character, leading to biased expectations concerning the AI-system's predictions. This might also be one factor explaining why people often believe the system to maximize their returns.

Finally, knowing that proposers have access to a system that is trained on prior interactions could cue evolutionary responder responses. Prior research has suggested that under the possibility that proposers have access to information about previous bargaining encounters, evolutionary dynamics might favor parity-demand strategies in order to protect the responder's reputation [127]. Here, the AI-system would function as a virtual data base triggering social signaling concerns. While these patterns would have developed within non-anonymous social groups, they might still affect behavior in anonymous social contexts.

RQ2: Are human preferences for or against AI-systems explained by economic expectations? Human avoidance of AI-systems cannot be explained by their economic expectations as quantified through incentivized belief elicitation. In fact, responders are even *more* likely to believe that the autonomous AI-system maximizes their expected income. While the majority of subjects who expect the human proposer to maximize their income subsequently also approach the human, only a small minority does so for the autonomous AI-system. For the AI decision aid, subjects are less likely to follow economic rationale than for the sole human, but more likely than for the autonomous AI-system. These results provide clear evidence that human avoidance of algorithmic AI agents is primarily a social phenomenon that overwrites predicted economic benefits. Further, the effect appears to be moderated by the system's autonomy. A deeper look into the data reveals that the existence of an autonomous AI-system potentially increases the acceptance of AI decision aids. While many subjects who expect maximal economic returns through the AI decision aid switch to the human, those who economically favor the autonomous AI-system spread relatively evenly across the other two types.

RQ3: Does the introduction of AI-systems into a bargaining environment affect overall economic welfare? The economic consequences of introducing AI-systems into existing bargaining environments hinges on three main factors: responder demands, proposer expectations, and responder self-selection. First, responders who bargain with an autonomous AI-system demand a larger share of the pie to successfully contract with the substituted human. This ceteris paribus unequivocally reduces the overall number of successful interactions and, thus, market efficiency. In our case, the effects were particularly damaging, as the system always offered the expected value maximizing offer calculated from mainly human-human interactions. On average, responders demanded more than what was offered, severely inhibiting the number of successful bargains. This highlights the maybe trivial insight that unforeseen changes in human behavior following the installation of AI-systems challenge their predictive accuracy and, thereby, usefulness. Second, proposers expected responders to increase economic demands when bargaining with the autonomous AI-system, and stated little inclination to rely on it themselves in a hypothetical scenario. Third, only a small minority of responders approached the autonomous AI-system. Thus, our results suggest that the introduction of algorithmic agents into bargaining environments has the potential to substantially hurt market outcomes, but is simultaneously unlikely to prevail if not carefully adjusted to a variety of human needs. A casually deployed AI-system will probably not crowd-out human competitors or even survive long enough to build reputation and exposure.

For the AI decision aid, the results are somewhat different. Responder demands do not differ from standard choices as observed throughout the literature and in this article. There is tangential evidence that proposers might expect responders to be a bit more assertive, but the differences are not significant. Additionally, most proposers favor being endowed with an AI decision aid. Thus, there appear to be fewer obstacles for AI-systems when deployed as a supporting system. One caveat is that many responders approaching the AI decision aid expected maximal economic returns from the autonomous AI-system. Without the latter, they might disproportionately opt for sole human proposers. Furthermore, given the same economic expectations, individuals or organizations deploying decision aids are still disadvantaged as they incur additional social costs.

5.1 Practical Implications

Understanding human reactions towards AI-systems and the underlying motivations behind their behavioral shifts is essential for a wide range of issues relevant to the broader HCI community. For one, effective system design and deployment requires a nuanced understanding of how people will receive and react to algorithmic agents [9, 120]. This is relevant both for systems to be *useful* and *usable*. Our results suggest that in bargaining environments, people might act more economically demanding towards AI-systems – a shift that disrupts known, predictable patterns of human behavior and thereby reduces the effectiveness of systems that rely on them to derive predictions. In theory, incorporating these changes into predictive models should mitigate losses. This, however, requires a comprehensive, nuanced and replicable codification of human decision making. We see our results as a first stepping stone towards that goal. Another interpretation of the data is that people who self-select into AI-system interactions without prior experience share some common characteristics that are not randomly distributed across the whole sample. In our experiment, responders who self-selected into bargaining were significantly more demanding. This highlights the problem of selection biases in machine-learning, as the system will be confronted with a particular set of beliefs, habits and demands that might not be representative of the wider population. The effectiveness of AI-systems applied in the real world

might therefore decrease with wider dissemination, as priorly aversive groups are increasingly exposed to algorithmic decisions that are calibrated towards different societal factions.

Another implication concerns the introduction of AI-tools into organizational bargaining environments. In particular, monetary incentives alone might not be enough to motivate widespread employee adoption. Delegating employee evaluation, remuneration or wage negotiation to algorithmic decision making systems could have adverse effects on worker behavior. On the consumer side, people might avoid service providers, professionals or sellers who noticeably rely on algorithmic decision units. This either disincentivizes AI-system usage, or promotes obfuscation. Borrowers receiving algorithmic credit scores could demand better conditions, c.p. reducing overall loaning volume, causing either a re-distribution of economic rents from financial institutes to individual borrowers through improved terms, or a negative shock to the overall number of successful loan applications. Thus, following the abstracted nature of this experiment, we argue that our results have explanatory value for a wide range of bargaining situations, while simultaneously recognizing that the different idiosyncrasies of various economic and societal domains might alleviate or exacerbate these effects.

In general, bargaining environments that resolve efficiently due to a combination of social and profit-seeking behavior [21, 153] could lose efficacy because of more assertive human choices and altered social expectations. Depending on the magnitude of the effect, this might necessitate additional institutional elements like monitoring or transparency rules. Regulators should be mindful of potentially adverse effects of increased AI-system dissemination on the overall number of successful market interactions.

More speculatively, our results also concern the design of incentive schemes and market mechanisms. Structures that combine social factors like reciprocity [58] or emotionally driven behavioral responses, such as altruistic and third-party punishment [55, 57], with strategic profit-seeking elements could cease to function with the introduction of AI-systems. Examples include contests and tournaments that determine group or individual wages [40, 65], efficiency wages [4] or team contributions and conditional cooperation [16, 97]. For mechanism design as well as the general construction of institutions, introducing AI-systems could disrupt existing models of human behavior. Many institutional frameworks concerning tax or trade schemes, voting and redistributive policies are based on social preferences like altruism, the assertiveness of ethical norms or intrinsic motivation [23, 28]. However, a lot of future research is needed to carefully parse-out the effect of AI-systems on human decision making conditional on the specific decision context. Whether the effects documented in this paper generalize to non-bargaining situations, or how various real-life variables such as transparency, obfuscation, interpretability, experience or interface design mediate human-algorithm relationships, remain open questions.

5.2 Caveats and Limitations

Our work has several limitations. For one, we do not provide a specific answer on what drives responder avoidance of the autonomous AI-system, or why people discount economic self-interest when faced with the possibility to actually bargain with an AI-system. While this article is the first to provide evidence regarding the negligence of economic motives, which might have been the most obvious explanation of behavioral adjustments, we rely on future research to parse out the different social motivations behind human behavior. Second, our work utilizes a very simple, synthesized model with little information for participants. It is very likely that human avoidance of AI-systems is mediated by a combination of interpretability and the specific kind of algorithm used [49, 52, 101, 141]. People have implicit priors and beliefs about AI-systems, and a systematic examination of how these expectations determine behavior, as well as how these expectations change with experience or through learning and information, could be worthwhile. For example, we did not disclose the system's specific objective function to responders. While they learned that the system "assisted" the human proposer or played "on behalf" of a passive human proposer, and therefore probably assumed it would aim to maximize proposer income, explicitly stating an algorithms goal might change behavior. While we would argue that most of the time, people either do not have access to the AI-system's objective function, or if they do, will almost never disclose it to the public, there might be instances where strategic transparency could be a viable organizational tool. Here, we rely on future research. Third, our experiments abstracts from many real-life challenges and contextual elements that could affect human behavior towards AI-systems. These include, but are not limited to: risk, uncertainty, task domain, privacy, prior experience, individual characteristics, habituation and familiarization, or the system's saliency. In reality, proposers might not negotiate over windfall gains, but their own earned money. While prior research suggests that windfall money (as well as stakes) does not substantially shift behavior in the ultimatum game for either role, small effects have sometimes been documented, which could be exacerbated beyond the lab [11, 15, 31, 103, 126]. In that case, we would expect more risk-averse proposer behavior [8, 26, 38, 131], which could in combination with the anticipated increase in responder demands further inhibit the supply-side dissemination of AI-systems. Rather than betting on a new technology with potentially adverse profit-effects, proposers would stick to the less uncertain, established option. Some preliminary evidence also points to responders accepting lower offers when proposers earn their income [105]. If true, proposers – e.g. sellers, employers – delegating decisions to AI-systems might be further 'punished' by foregoing their own effort and thereby crowding-out responder recognition of their work. Finally, in our experiment, the system was a defining element, potentially over-stating the effect for situations where people are already used to relying on algorithms. For example, most humans do not seem to have a problem with using AI technologies via search engines, for navigation, or in computer games. Instead, our results might be more suitable for emerging systems in domains that either until recently have been, or still are dominated by purely human interactions.

6 CONCLUSION

This article leverages an extended ultimatum game to gauge human preferences for AI-systems with different levels of autonomy and quantify subsequent bargaining behavior. Further, we exploit incentivized belief elicitation to map behavior to economic expectations

and thus distinguish between economic and social motivations. Our results demonstrate that responders overwhelmingly self-select into bargaining with a human proposer at the expense of an autonomous AI-system that decides on behalf of another human. These preferences often contradict subjects' economic expectations, highlighting how people overwrite economic self-interest in order to avoid bargaining with AI-systems. Responders are most likely to maximize their expected income when it coincides with approaching a sole human proposer. Introducing AI-systems with progressively more decision autonomy decreases maximization behavior. Effects are by far the most pronounced for the fully autonomous AI-system. Further, responders who decide to bargain with an autonomous AI-system are significantly more demanding, hurting overall economic welfare. Proposers expect these behavioral changes, and state low willingness to rely on an autonomous system themselves. Overall, there seem to be substantial social obstacles for AI-systems to compete with human actors in bargaining environments, as people are less motivated by economic rationale while exhibiting strong preferences to avoid AI-systems.

ACKNOWLEDGMENTS

We would like to thank the anonymous participants in our study and the reviewers who provided valuable feedback. This work was supported by the Lower Saxony Ministry of Science and Culture under grant number ZN3492 within the Lower Saxony "Vorab" of the Volkswagen Foundation, the Center for Digital Innovations (ZDIN), and the TU Delft Design@Scale AI lab within the TU Delft AI Initiative.

REFERENCES

[1] Marc TP Adam, Jan Krämer, and Marius B Müller. 2015. Auction fever! How time pressure and social competition affect bidders' arousal and bids in retail auctions. *Journal of Retailing* 91, 3 (2015), 468–485.
[2] Marc TP Adam, Timm Teubner, and Henner Gimpel. 2018. No rage against the machine: how computer agents mitigate human emotional processes in electronic negotiations. *Group Decision and Negotiation* 27, 4 (2018), 543–571.
[3] Eyal Aharoni and Alan J Fridlund. 2007. Social reactions toward people vs. computers: How mere lables shape interactions. *Computers in human behavior* 23, 5 (2007), 2175–2189.
[4] George A Akerlof. 1984. Gift exchange and efficiency-wage theory: Four views. *The American Economic Review* 74, 2 (1984), 79–83.
[5] James Andreoni. 1990. Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal* 100, 401 (1990), 464–477.
[6] James Andreoni and B Douglas Bernheim. 2009. Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica* 77, 5 (2009), 1607–1636.
[7] Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. 2020. In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY* 35, 3 (2020), 611–623.
[8] Hal R Arkes, Cynthia A Joyner, Mark V Pezzo, Jane Gradwohl Nash, Karen Siegel-Jacobs, and Eric Stone. 1994. The psychology of windfall gains. *Organizational behavior and human decision processes* 59, 3 (1994), 331–347.
[9] Liam J Bannon. 1995. From human factors to human actors: The role of psychology and human-computer interaction studies in system design. In *Readings in human–computer interaction*. Elsevier, 205–214.
[10] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S Weld, Walter S Lasecki, and Eric Horvitz. 2019. Updates in human-ai teams: Understanding and addressing the performance/compatibility tradeoff. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 2429–2437.
[11] Benjamin S Barber IV and William English. 2019. The origin of wealth matters: Equity norms trump equality norms in the ultimatum game with earned endowments. *Journal of Economic Behavior & Organization* 158 (2019), 33–43.
[12] Christoph Bartneck, Kumar Yogeeswaran, Qi Min Ser, Graeme Woodward, Robert Sparrow, Siheng Wang, and Friederike Eyssel. 2018. Robots and racism. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 196–204.
[13] Volker Benndorf, Thomas Große Brinkhaus, and Ferdinand von Siemens. 2021. Ultimatum Game Behavior in a Social-Preferences Vacuum Chamber. (2021).
[14] Matthias Benz and Stephan Meier. 2008. Do people behave in experiments as in the field?—evidence from donations. *Experimental economics* 11, 3 (2008), 268–281.
[15] Roger Berger, Heiko Rauhut, Sandra Prade, and Dirk Helbing. 2012. Bargaining over waiting time in ultimatum game experiments. *Social science research* 41, 2 (2012), 372–379.
[16] Theodore Bergstrom, Lawrence Blume, and Hal Varian. 1986. On the private provision of public goods. *Journal of public economics* 29, 1 (1986), 25–49.
[17] Maik Bieleke, Peter M Gollwitzer, Gabriele Oettingen, and Urs Fischbacher. 2017. Social value orientation moderates the effects of intuition versus reflection on responses to unfair ultimatum offers. *Journal of Behavioral Decision Making* 30, 2 (2017), 569–581.
[18] Yochanan E Bigman and Kurt Gray. 2018. People are averse to machines making moral decisions. *Cognition* 181 (2018), 21–34.
[19] Ekkehart Boehmer, Kingsley Fong, and Julie Wu. 2012. International evidence on algorithmic trading. In *AFA 2013 San Diego Meetings Paper*.
[20] Miranda Bogen and Aaron Rieke. 2018. Help wanted: An examination of hiring algorithms, equity, and bias. (2018).
[21] Iris Bohnet and Richard Zeckhauser. 2004. Social comparisons in ultimatum bargaining. *Scandinavian Journal of Economics* 106, 3 (2004), 495–510.
[22] Alessandro Bonatti and Gonzalo Cisternas. 2020. Consumer scores and price discrimination. *The Review of Economic Studies* 87, 2 (2020), 750–791.
[23] Samuel Bowles and Sung-Ha Hwang. 2008. Social preferences and public economics: Mechanism design when social preferences depend on incentives. *Journal of public economics* 92, 8-9 (2008), 1811–1820.
[24] Hermann Brandstätter and Manfred Königstein. 2001. Personality influences on ultimatum bargaining decisions. *European Journal of Personality* 15, S1 (2001), S53–S70.
[25] Kristin M Brethel-Haurwitz, Sarah A Stoycos, Elise M Cardinale, Bryce Huebner, and Abigail A Marsh. 2016. Is costly punishment altruistic? Exploring rejection of unfair offers in the Ultimatum Game in real-world altruists. *Scientific reports* 6, 1 (2016), 1–10.
[26] Joseph Briggs, David Cesarini, Erik Lindqvist, and Robert Östling. 2021. Windfall gains and stock market participation. *Journal of Financial Economics* 139, 1 (2021), 57–83.
[27] Zach Y Brown and Alexander MacKay. 2021. *Competition in pricing algorithms*. Technical Report. National Bureau of Economic Research.
[28] Antonio Cabrales, Raffaele Miniaci, Marco Piovesan, and Giovanni Ponti. 2010. Social preferences and strategic uncertainty: an experiment on markets and contracts. *American Economic Review* 100, 5 (2010), 2261–78.
[29] Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolo, and Sergio Pastorello. 2020. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review* 110, 10 (2020), 3267–97.
[30] Colin Camerer. 2011. The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. *Available at SSRN 1977749* (2011).
[31] Alexander W Cappelen, Trond Halvorsen, Erik Ø Sørensen, and Bertil Tungodden. 2017. Face-saving or fair-minded: What motivates moral behavior? *Journal of the European Economic Association* 15, 3 (2017), 540–557.
[32] John Cartlidge, Marco De Luca, Charlotte Szostek, and Dave Cliff. 2012. Too fast too furious: faster financial-market trading agents can give less efficient markets. In *ICAART-2012: 4th International Conference on Agents and Artificial Intelligence*. SciTePress, 126–135.
[33] Noah Castelo, Maarten W Bos, and Donald R Lehmann. 2019. Task-dependent algorithm aversion. *Journal of Marketing Research* 56, 5 (2019), 809–825.
[34] Thierry Chaminade, Delphine Rosset, David Da Fonseca, Bruno Nazarian, Ewald Lutscher, Gordon Cheng, and Christine Deruelle. 2012. How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in human neuroscience* 6 (2012), 103.
[35] Gary Charness and Ernst Fehr. 2015. From the lab to the real world. *Science* 350, 6260 (2015), 512–513.
[36] Le Chen, Alan Mislove, and Christo Wilson. 2016. An empirical analysis of algorithmic pricing on amazon marketplace. In *Proceedings of the 25th international conference on World Wide Web*. 1339–1349.
[37] Le Chen and Christo Wilson. 2017. Observing algorithmic marketplaces in-the-wild. *ACM SIGecom Exchanges* 15, 2 (2017), 34–39.
[38] Jeremy Clark. 2002. House money effects in public good experiments. *Experimental Economics* 5, 3 (2002), 223–231.
[39] Alain Cohn, Tobias Gesche, and Michel André Maréchal. 2020. *Honesty in the digital age*. Technical Report. Management Science (forthcoming).
[40] Brian L Connelly, Laszlo Tihanyi, T Russell Crook, and K Ashley Gangloff. 2014. Tournament theory: Thirty years of contests and competitions. *Journal of Management* 40, 1 (2014), 16–47.
[41] Jacob W Crandall, Mayada Oudah, Fatimah Ishowo-Oloko, Sherief Abdallah, Jean-François Bonnefon, Manuel Cebrian, Azim Shariff, Michael A Goodrich, Iyad Rahwan, et al. 2018. Cooperating with machines. *Nature communications*

9, 1 (2018), 1–12.
[42] Maria De-Arteaga, Riccardo Fogliato, and Alexandra Chouldechova. 2020. A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[43] Marco De Luca and Dave Cliff. 2011. Human-agent auction interactions: Adaptive-aggressive agents dominate. In *Twenty-second international joint conference on artificial intelligence*.
[44] Celso M de Melo, Jonathan Gratch, and Peter J Carnevale. 2014. Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing* 6, 2 (2014), 127–136.
[45] Celso M de Melo, Peter Khooshabeh, Ori Amir, and Jonathan Gratch. 2018. Shaping cooperation between humans and agents with emotion expressions and framing. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 2224–2226.
[46] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144, 1 (2015), 114.
[47] Berkeley J Dietvorst, Joseph P Simmons, and Cade Massey. 2018. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 64, 3 (2018), 1155–1170.
[48] Nicole Dillen, Marko Ilievski, Edith Law, Lennart E Nacke, Krzysztof Czarnecki, and Oliver Schneider. 2020. Keep calm and ride along: Passenger comfort and anxiety as physiological responses to autonomous driving styles. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[49] Finale Doshi-Velez and Been Kim. 2017. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).
[50] Christoph Eisenegger, Michael Naef, Romana Snozzi, Markus Heinrichs, and Ernst Fehr. 2010. Prejudice and truth about the effect of testosterone on human bargaining behaviour. *Nature* 463, 7279 (2010), 356–359.
[51] Matthew Embrey, Guillaume R Fréchette, and Steven F Lehrer. 2015. Bargaining and reputation: An experiment on bargaining in the presence of behavioural types. *The Review of Economic Studies* 82, 2 (2015), 608–631.
[52] Alexander Erlei, Franck Nekdem, Lukas Meub, Avishek Anand, and Ujwal Gadiraju. 2020. Impact of Algorithmic Decision Making on Human Behavior: Evidence from Ultimatum Bargaining. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. 43–52.
[53] Armin Falk. 2007. Gift exchange in the field. *Econometrica* 75, 5 (2007), 1501–1511.
[54] Mike Farjam and Oliver Kirchkamp. 2018. Bubbles in hybrid markets: How expectations about algorithmic trading affect human trading. *Journal of Economic Behavior & Organization* 146 (2018), 248–269.
[55] Ernst Fehr and Urs Fischbacher. 2004. Third-party punishment and social norms. *Evolution and human behavior* 25, 2 (2004), 63–87.
[56] Ernst Fehr and Simon Gächter. 2000. Fairness and retaliation: The economics of reciprocity. *Journal of economic perspectives* 14, 3 (2000), 159–181.
[57] Ernst Fehr and Simon Gächter. 2002. Altruistic punishment in humans. *Nature* 415, 6868 (2002), 137–140.
[58] Ernst Fehr, Erich Kirchler, Andreas Weichbold, and Simon Gächter. 1998. When social norms overpower competition: Gift exchange in experimental labor markets. *Journal of Labor economics* 16, 2 (1998), 324–351.
[59] BJ Fogg and Clifford Nass. 1997. How users reciprocate to computers: an experiment that demonstrates behavior change. In *CHI'97 extended abstracts on Human factors in computing systems*. 331–332.
[60] Axel Franzen and Sonja Pointner. 2013. The external validity of giving in the dictator game. *Experimental Economics* 16, 2 (2013), 155–169.
[61] Ujwal Gadiraju, Ricardo Kawase, Stefan Dietze, and Gianluca Demartini. 2015. Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1631–1640.
[62] Uri Gneezy and John A List. 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74, 5 (2006), 1365–1384.
[63] Jan Gogoll and Matthias Uhl. 2018. Rage against the machine: Automation in the moral domain. *Journal of Behavioral and Experimental Economics* 74 (2018), 97–103.
[64] Binglin Gong and Chun-Lei Yang. 2012. Gender differences in risk attitudes: Field experiments on the matrilineal Mosuo and the patriarchal Yi. *Journal of economic behavior & organization* 83, 1 (2012), 59–65.
[65] Anna Gunnthorsdottir and Amnon Rapoport. 2006. Embedding social dilemmas in intergroup competition reduces free-riding. *Organizational Behavior and Human Decision Processes* 101, 2 (2006), 184–199.
[66] Akshit Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. 2022. To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System. (2022).
[67] Werner Güth and Martin G Kocher. 2014. More than thirty years of ultimatum bargaining experiments: Motives, variations, and a survey of the recent literature. *Journal of Economic Behavior & Organization* 108 (2014), 396–409.
[68] Werner Güth, Carsten Schmidt, and Matthias Sutter. 2007. Bargaining outside the lab–a newspaper experiment of a three-person ultimatum game. *The Economic Journal* 117, 518 (2007), 449–469.
[69] Werner Güth, Rolf Schmittberger, and Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization* 3, 4 (1982), 367–388.
[70] Tessa Haesevoets, David De Cremer, Kim Dierckx, and Alain Van Hiel. 2021. Human-machine collaboration in managerial decision making. *Computers in Human Behavior* 119 (2021), 106730.
[71] Jens Hainmueller, Dominik Hangartner, and Teppei Yamamoto. 2015. Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences* 112, 8 (2015), 2395–2400.
[72] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*. 305–318.
[73] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. 2017. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 1914–1933.
[74] Terrence Hendershott, Charles M Jones, and Albert J Menkveld. 2011. Does algorithmic trading improve liquidity? *The Journal of finance* 66, 1 (2011), 1–33.
[75] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. *American Economic Review* 91, 2 (2001), 73–78.
[76] Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, et al. 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences* 28, 6 (2005), 795–815.
[77] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences* 33, 2-3 (2010), 61–83.
[78] Daniel Herbst and Alexandre Mas. 2015. Peer effects on worker output in the laboratory generalize to the field. *Science* 350, 6260 (2015), 545–549.
[79] Nicholas Hertz and Eva Wiese. 2016. Influence of agent type and task ambiguity on conformity in social decision making. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 313–317.
[80] Arend Hintze and Ralph Hertwig. 2016. The evolution of generosity in the ultimatum game. *Scientific reports* 6, 1 (2016), 1–7.
[81] Tom Hitron, Yoav Orlev, Iddo Wald, Ariel Shamir, Hadas Erel, and Oren Zuckerman. 2019. Can children understand machine learning concepts? The effect of uncovering black boxes. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–11.
[82] Tanjim Hossain and Ryo Okui. 2013. The binarized scoring rule. *Review of Economic Studies* 80, 3 (2013), 984–1001.
[83] John W Huppertz, Sidney J Arenson, and Richard H Evans. 1978. An application of equity theory to buyer-seller exchange situations. *Journal of marketing research* 15, 2 (1978), 250–260.
[84] Fatimah Ishowo-Oloko, Jean-François Bonnefon, Zakariyah Soroye, Jacob Crandall, Iyad Rahwan, and Talal Rahwan. 2019. Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nature Machine Intelligence* 1, 11 (2019), 517–521.
[85] Shayan Jawed, Ahmed Rashed, Kiran Madhusudhanan, Shereen Elsayed, Mohsan Jameel, Alexei Volk, Andre Hintsches, Marlies Kornfeld, Katrin Lange, Lars Schmidt-Thieme, et al. 2022. AI and Data-Driven Mobility at Volkswagen Financial Services AG. *arXiv preprint arXiv:2202.04411* (2022).
[86] Eric J Johnson, Colin Camerer, Sankar Sen, and Talia Rymon. 2002. Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory* 104, 1 (2002), 16–47.
[87] Dominik Jung, Verena Dorner, Florian Glaser, and Stefan Morana. 2018. Robo-advisory. *Business & Information Systems Engineering* 60, 1 (2018), 81–86.
[88] Tanja B Jutzi, Eva I Krieghoff-Henning, Tim Holland-Letz, Jochen Sven Utikal, Axel Hauschild, Dirk Schadendorf, Wiebke Sondermann, Stefan Fröhling, Achim Hekler, Max Schmitt, et al. 2020. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Frontiers in medicine* 7 (2020), 233.
[89] Daniel Kahneman, Jack L Knetsch, and Richard Thaler. 1986. Fairness as a constraint on profit seeking: Entitlements in the market. *The American economic review* (1986), 728–741.
[90] Laura Kaltwasser, Andrea Hildebrandt, Oliver Wilhelm, and Werner Sommer. 2016. Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. *Social Cognitive and Affective Neuroscience* 11, 10 (2016), 1608–1617.
[91] Gokhan Karagonlar and David M Kuhlman. 2013. The role of social value orientation in response to an unfair offer in the ultimatum game. *Organizational Behavior and Human Decision Processes* 120, 2 (2013), 228–239.

[92] Dean S Karlan. 2005. Using experimental economics to measure social capital and predict financial decisions. *American Economic Review* 95, 5 (2005), 1688–1699.

[93] Yasuhiro Katagiri, Clifford Nass, and Yugo Takeuchi. 2001. Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (2001), 1558–1562.

[94] Navroop Kaur and Sandeep K Sood. 2015. A game theoretic approach for an IoT-based automated employee performance evaluation. *IEEE Systems Journal* 11, 3 (2015), 1385–1394.

[95] Katherine C Kellogg, Melissa A Valentine, and Angele Christin. 2020. Algorithms at work: The new contested terrain of control. *Academy of Management Annals* 14, 1 (2020), 366–410.

[96] Rudolf Kerschbamer, Daniel Neururer, and Matthias Sutter. 2019. Credence goods markets and the informational value of new media: A natural field experiment. *MPI collective goods discussion paper* 2019/3 (2019).

[97] Claudia Keser and Frans Van Winden. 2000. Conditional cooperation and voluntary contributions to public goods. *scandinavian Journal of Economics* 102, 1 (2000), 23–39.

[98] Carrie L Kovarik. 2020. Patient perspectives on the use of artificial intelligence. *JAMA dermatology* 156, 5 (2020), 493–494.

[99] Max F Kramer, Jana Schaich Borg, Vincent Conitzer, and Walter Sinnott-Armstrong. 2018. When do people want AI to make decisions?. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 204–209.

[100] Nathan R Kuncel, David M Klieger, and Deniz S Ones. 2014. In hiring, algorithms beat instinct. *Harvard business review* 92, 5 (2014), p32–32.

[101] Isaac Lage, Andrew Slavin Ross, Been Kim, Samuel J Gershman, and Finale Doshi-Velez. 2018. Human-in-the-loop interpretability prior. *Advances in neural information processing systems* 31 (2018).

[102] Markus Langer, Cornelius J König, and Maria Papathanasiou. 2019. Highly automated job interviews: Acceptance under the influence of stakes. *International Journal of Selection and Assessment* 27, 3 (2019), 217–234.

[103] Andrea Larney, Amanda Rotella, and Pat Barclay. 2019. Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes* 151 (2019), 61–72.

[104] Susan K Laury and Laura O Taylor. 2008. Altruism spillovers: Are behaviors in context-free experiments predictive of altruism toward a naturally occurring public good? *Journal of Economic Behavior & Organization* 65, 1 (2008), 9–29.

[105] Kangoh Lee and Quazi Shahriar. 2017. Fairness, One's Source of Income, and Others' Decisions: An Ultimatum Game Experiment. *Managerial and Decision Economics* 38, 3 (2017), 423–431.

[106] Min Kyung Lee. 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society* 5, 1 (2018), 2053951718756684.

[107] Tian-Shyug Lee and I-Fei Chen. 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with applications* 28, 4 (2005), 743–752.

[108] Steven D Levitt and John A List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives* 21, 2 (2007), 153–174.

[109] Danielle Li, Lindsey R Raymond, and Peter Bergman. 2020. *Hiring as exploration*. Technical Report. National Bureau of Economic Research.

[110] Yang Li, Wanshan Zheng, and Zibin Zheng. 2019. Deep robust reinforcement learning for practical algorithmic trading. *IEEE Access* 7 (2019), 108014–108022.

[111] Yuan-shuh Lii and Erin Sy. 2009. Internet differential pricing: Effects on consumer price perception, emotions, and behavioral responses. *Computers in Human Behavior* 25, 3 (2009), 770–777.

[112] Sohye Lim and Byron Reeves. 2010. Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies* 68, 1-2 (2010), 57–68.

[113] Mai Skjott Linneberg and Steffen Korsgaard. 2019. Coding qualitative data: A synthesis guiding the novice. *Qualitative research journal* (2019).

[114] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[115] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–16.

[116] Chiara Longoni, Andrea Bonezzi, and Carey K Morewedge. 2019. Resistance to medical artificial intelligence. *Journal of Consumer Research* 46, 4 (2019), 629–650.

[117] Zhuoran Lu and Ming Yin. 2021. Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.

[118] Lydia Manikonda and Subbarao Kambhampati. 2018. Tweeting AI: perceptions of lay versus expert twitterati. In *Twelfth International AAAI Conference on Web and Social Media*.

[119] Christoph March. 2021. Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology* (2021), 102426.

[120] Elisa D Mekler and Kasper Hornbæk. 2019. A framework for the experience of meaning in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[121] Celso De Melo, Stacy Marsella, and Jonathan Gratch. 2016. People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 2 (2016), 1–17.

[122] Ryan O Murphy, Kurt A Ackermann, and Michel Handgraaf. 2011. Measuring social value orientation. *Judgment and Decision making* 6, 8 (2011), 771–781.

[123] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[124] Andreas Nicklisch and Irenaeus Wolff. 2012. On the nature of reciprocity: Evidence from the ultimatum reciprocity measure. *Journal of Economic Behavior & Organization* 84, 3 (2012), 892–905.

[125] Paweł Niszczota and Dániel Kaszás. 2020. Robo-investment aversion. *Plos one* 15, 9 (2020), e0239277.

[126] Charles N Noussair and Jan Stoop. 2015. Time as a medium of reward in three social preference experiments. *Experimental Economics* 18, 3 (2015), 442–456.

[127] Martin A Nowak, Karen M Page, and Karl Sigmund. 2000. Fairness versus reason in the ultimatum game. *Science* 289, 5485 (2000), 1773–1775.

[128] Christian J Park, H Yi Paul, and Eliot L Siegel. 2021. Medical student perspectives on the impact of artificial intelligence on the practice of medicine. *Current problems in diagnostic radiology* 50, 5 (2021), 614–619.

[129] Hyanghee Park, Daehwan Ahn, Kartik Hosanagar, and Joonhwan Lee. 2021. Human-AI Interaction in Human Resource Management: Understanding Why Employees Resist Algorithmic Evaluation at Workplaces and How to Mitigate Burdens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[130] David C Parkes and Michael P Wellman. 2015. Economic reasoning and artificial intelligence. *Science* 349, 6245 (2015), 267–272.

[131] Jiaxi Peng, Danmin Miao, and Wei Xiao. 2013. Why are gainers more risk seeking. *Judgment & Decision Making* 8, 2 (2013).

[132] Marco Perugini, Marcello Gallucci, Fabio Presaghi, and Anna Paola Ercolani. 2003. The personal norm of reciprocity. *European Journal of Personality* 17, 4 (2003), 251–283.

[133] Maria Petrescu and Dhruv Bhatli. 2013. Consumer behavior in flea markets and marketing to the Bottom of the Pyramid. *Journal of Management Research* 13, 1 (2013), 55–63.

[134] Anders U Poulsen and Jonathan HW Tan. 2007. Information acquisition in the ultimatum game: An experimental study. *Experimental Economics* 10, 4 (2007), 391–409.

[135] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. 2019. Machine behaviour. *Nature* 568, 7753 (2019), 477–486.

[136] David G Rand, Corina E Tarnita, Hisashi Ohtsuki, and Martin A Nowak. 2013. Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proceedings of the National Academy of Sciences* 110, 7 (2013), 2581–2586.

[137] John R Rossiter and Alvin M Chan. 1998. Ethnicity in business and consumer behavior. *Journal of Business Research* 42, 2 (1998), 127–134.

[138] Kasper Roszbach. 2004. Bank lending policy, credit scoring, and the survival of loans. *Review of Economics and Statistics* 86, 4 (2004), 946–958.

[139] Nicole Salomons, Michael Van Der Linden, Sarah Strohkorb Sebo, and Brian Scassellati. 2018. Humans conform to robots: Disambiguating trust, truth, and conformity. In *Proceedings of the 2018 acm/ieee international conference on human-robot interaction*. 187–195.

[140] Alan G Sanfey, James K Rilling, Jessica A Aronson, Leigh E Nystrom, and Jonathan D Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300, 5626 (2003), 1755–1758.

[141] Philipp Schmidt and Felix Biessmann. 2019. Quantifying interpretability and trust in machine learning systems. *arXiv preprint arXiv:1901.08558* (2019).

[142] Eric Schniter, Timothy W Shields, and Daniel Sznycer. 2020. Trust in humans and robots: Economically similar but emotionally different. *Journal of Economic Psychology* 78 (2020), 102253.

[143] Peter Seele, Claus Dierksmeier, Reto Hofstetter, and Mario D Schultz. 2021. Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. *Journal of Business Ethics* 170, 4 (2021), 697–719.

[144] Donghee Shin, Bouziane Zaid, and Mohammed Ibahrine. 2020. Algorithm Appreciation: Algorithmic Performance, Developmental Processes, and User Interactions. In *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*. IEEE, 1–5.

[145] Bethany Stai, Nick Heller, Sean McSweeney, Jack Rickman, Paul Blake, Ranveer Vasdev, Zach Edgerton, Resha Tejpaul, Matt Peterson, Joel Rosenberg, et al. 2020. Public perceptions of artificial intelligence and robotics in medicine. *Journal of endourology* 34, 10 (2020), 1041–1048.

[146] Tuomas Takko, Kunal Bhattacharya, Daniel Monsivais, and Kimmo Kaski. 2021. Human-agent coordination in a group formation game. *Scientific Reports* 11, 1 (2021), 1–10.
[147] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior* 38 (2014), 75–84.
[148] Timm Teubner, Marc Adam, and Ryan Riordan. 2015. The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions. *Journal of the Association for Information Systems* 16, 10 (2015), 2.
[149] Suzanne Tolmeijer, Ujwal Gadiraju, Ramya Ghantasala, Akshit Gupta, and Abraham Bernstein. 2021. Second chance for a first impression? Trust development in intelligent system interaction. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. 77–87.
[150] Stefan T Trautmann and Gijs van de Kuilen. 2015. Belief elicitation: A horse race among truth serums. *The Economic Journal* 125, 589 (2015), 2116–2135.
[151] Chih-Fong Tsai and Jhen-Wei Wu. 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications* 34, 4 (2008), 2639–2649.
[152] Eric Van Damme, Kenneth G Binmore, Alvin E Roth, Larry Samuelson, Eyal Winter, Gary E Bolton, Axel Ockenfels, Martin Dufwenberg, Georg Kirchsteiger, Uri Gneezy, et al. 2014. How Werner Güth's ultimatum game shaped our understanding of social behavior. *Journal of economic behavior & organization* 108 (2014), 292–318.
[153] Eric Van Dijk, David De Cremer, and Michel JJ Handgraaf. 2004. Social value orientations and the strategic use of fairness in ultimatum bargaining. *Journal of experimental social psychology* 40, 6 (2004), 697–707.
[154] Mascha Van't Wout, René S Kahn, Alan G Sanfey, and André Aleman. 2006. Affective state and decision-making in the ultimatum game. *Experimental brain research* 169, 4 (2006), 564–568.
[155] Mascha Van't Wout, René S Kahn, Alan G Sanfey, and André Aleman. 2006. Affective state and decision-making in the ultimatum game. *Experimental brain research* 169, 4 (2006), 564–568.
[156] Arjan Verschoor, Ben D'Exelle, and Borja Perez-Viana. 2016. Lab and life: Does risky choice behaviour observed in experiments reflect that in the real world? *Journal of Economic Behavior & Organization* 128 (2016), 134–148.
[157] Benjamin von Walter, Dietmar Kremmel, and Bruno Jäger. 2021. The impact of lay beliefs about AI on adoption of algorithmic advice. *Marketing Letters* (2021), 1–13.
[158] Jeffrey Warshaw, Tara Matthews, Steve Whittaker, Chris Kau, Mateo Bengualid, and Barton A Smith. 2015. Can an Algorithm Know the" Real You"? Understanding People's Reactions to Hyper-personal Analytics Systems. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 797–806.
[159] Lan Xia, Kent B Monroe, and Jennifer L Cox. 2004. The price is unfair! A conceptual framework of price fairness perceptions. *Journal of marketing* 68, 4 (2004), 1–15.
[160] Toshio Yamagishi, Yutaka Horita, Nobuhiro Mifune, Hirofumi Hashimoto, Yang Li, Mizuho Shinada, Arisa Miura, Keigo Inukai, Haruto Takagishi, and Dora Simunovic. 2012. Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences* 109, 50 (2012), 20364–20368.
[161] Takafumi Yamakawa, Yoshitaka Okano, and Tatsuyoshi Saijo. 2016. Detecting motives for cooperation in public goods experiments. *Experimental Economics* 19, 2 (2016), 500–512.
[162] April D Young and Andrew E Monroe. 2019. Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology* 85 (2019), 103870.
[163] Niklas Zethraeus, Ljiljana Kocoska-Maras, Tore Ellingsen, BO Von Schoultz, Angelica Linden Hirschberg, and Magnus Johannesson. 2009. A randomized trial of the effect of estrogen and testosterone on economic behavior. *Proceedings of the National Academy of Sciences* 106, 16 (2009), 6535–6538.
[164] Wenyu Zhang, Hongliang He, and Shuai Zhang. 2019. A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. *Expert Systems with Applications* 121 (2019), 221–232.
[165] Zhan Zhang, Yegin Genc, Aiwen Xing, Dakuo Wang, Xiangmin Fan, and Daniel Citardi. 2020. Lay individuals' perceptions of artificial intelligence (AI)-empowered healthcare systems. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e326.
[166] Kun Zhao and Luke D Smillie. 2015. The role of interpersonal traits in social decision making: Exploring sources of behavioral heterogeneity in economic games. *Personality and Social Psychology Review* 19, 3 (2015), 277–302.
[167] Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C Parkes, and Richard Socher. 2020. The ai economist: Improving equality and productivity with ai-driven tax policies. *arXiv preprint arXiv:2004.13332* (2020).
[168] Jichen Zhu, Jennifer Villareale, Nithesh Javvaji, Sebastian Risi, Mathias Löwe, Rush Weigelt, and Casper Harteveld. 2021. Player-AI Interaction: What Neural Network Games Reveal About AI as Play. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.