Strobel, Christina

**Conference Paper**

# The Hidden Costs of Automation

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics

**Provided in Cooperation with:**
Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Strobel, Christina (2022) : The Hidden Costs of Automation, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2022: Big Data in Economics, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/264129

# The hidden costs of automation*

Christina Strobel†

March 1, 2022

Business Process Automation is more and more replacing humans in management processes. The aim of this paper is to examine whether the use of such automated processes influences human performance and whether it matters who decides to use such an automated process. By applying a modified Principal-Agent Game ran on the microtask marketplace Amazon Mechanical Turk (Amazon MTurk), I compare performance when using an *automated Performance Appraisal System* (*automated PAS*) versus a non-automated, *manual Performance Appraisal System* (*manual PAS*). Depending on the treatment, either an individual or a random mechanism decides what type of system is going to be used. I find that performance is significantly lower in an automated system than in a non-automated system. However, performance does not differ significantly depending on whether an individual or a random mechanism decides to use the automated system or not.

*Keywords*: Principal-Agent-Setting; Automation; Performance-related Pay; Performance; Intelligent Process Automation
*JEL classification*: C91, D63, D80

*This document was created on March 1, 2022, with R version 3.4.1 (2017-06-30), on x86_64-w64-mingw32.

†Hamburg University of Technology, Institute for Digital Economics, Blohmstraße 15, 21079 Hamburg, Christina.Strobel@tuhh.de.

# 1. Introduction

Algorithm-controlled systems take over more and more tasks from humans. This is not only the case in manufacturing, where manufacturing robots replace assembly line workers, but also in management, where the level of process automation continues to increase.[1]

As part of the digital transformation, management processes are being simplified, streamlined and continuously automated. The technology-enabled automation of complex business processes and tasks is called Intelligent Process Automation (IPA) (Chakraborti et al., 2020). IPA extends across a wide range of business areas such as marketing, sales, procurement and workforce management (Manyika et al., 2017). Automation is supported by different software solutions, leveraging machine learning and Artificial Intelligence (AI) methods (Mohanty and Vyas, 2018).

One area in which many automated processes are already widespread as of today is human resources. According to a study by the international consultancy firm PricewaterhouseCoopers, 40% of the HR-functions in international companies are already using AI-based automation (Charlier and Kloppenburg, 2017). Typical tasks for such algorithms are supporting with the selection of future employees by scoring and selecting the most suited applicants (Upadhyay and Khandelwal, 2018). Another example is the use of automated processes to evaluate performance.[2] Here, automated processes are replacing direct human-to-human interactions in employee performance evaluations, e.g. to determine bonus payments (Kaur and Sood, 2017). While in the past a supervisor would decide about a final bonus payment, bonus assessments are becoming more and more automated. Software solutions based on semantic models allow to collect and evaluate data and determine job performance appraisal decisions with minimal human involvement (Yen et al., 2017). Due to this, an employee's bonus payment depends much less on an individual's assessment than on a predetermined algorithm.

While such automation might increase equality, as it e.g. reduces personal (potentialy biased) impressions, little attention has been paid to possible unintended effects of the reduction of the human factor. This leads to questions such as: how do people perceive and react to automated decisions in management processes?

Former research has shown that work performance is influenced by situational circumstances.[3] Letting an algorithm instead of a human decide means changing the situational circumstances of the underlying process. Hence, when replacing human-to-human interactions with machines, robots and automated processes, it can be assumed that these technologies also have an impact on performance.

Another influencing factor might be whether the direct supervisor or the company in general decides to use an automated approach to evaluate the performance. We know from former research by Fehr et al. (1993) that people tend to reward kind actions and to punish

[1] Forecasts predict a compound annual growth rate of 10.5% for the Business Process Management Industry till 2025 (Markets and Markets, 2020).

[2] Performance and effort provision are used interchangeable in the paper.

[3] Falk and Kosfeld (2006) found that setting a minimum performance requirement has a negative effect on performance.

unkind ones in an experimental labor market.[4]

If implementing a more automated approach is interpreted as a decrease in appreciation and/or trust, the use of such a system might lead to a decrease in the employee's satisfaction and, in turn, work performance. Thus, when an individual decides to automate a management process, workers might respond to the decision either by showing a higher or lower performence, depending on their perception of the decision. However, if the company as a superior entity decides about whether the process is going to be automate a process or not, reciprocal behavior can be expected to be less pronounced due to the lack of an identifiable counterparty.

Using an algorithm generally makes a process less individualizable and more rigid (Stone, 1971). Given the rapid development of new, scalable IPA solutions, it is inevitable to investigate how people react to such process automation. In this paper, I shed light on the influence process automation in management has on performance, and investigate whether it matters who made the decision to automate: the supervior (second-party control) or the company (third-party control). In particular, I focus on process automation in the bonus payment process.

The remainder of the paper is organized as follows: Section 2 provides a literature review focusing on experimental evidence from economics and social psychology research. In Section 3, I describe the basic experimental design. Then, in Section 4, I relate the experiment to the theoretical background and derive behavioral predictions. I present the results in Section 5. Section 6 concludes the paper by summarizing the main findings and discussing their implications as well as identifying further research ideas.

## 2. Related literature

The effect of bonus payments on performance has been researched extensively. Fehr and Schmidt (2007) show that a bonus contract offered by the principal in a chosen-effort Principal-Agent experiment leads to higher effort provision by the agent than a contract that fines the agent, or a trust contract where the principal offers a fixed wage. Fehr et al. (2007) confirm this observation by finding bonus contracts that rely on fairness and trust as an enforcement device to be more efficient and more profitable than incentive contracts enforced by the courts. In their experiment, the principal was able to choose a mechanism to enforce a specific effort from the agent with the support of a third party or to announce a non-binding, voluntary bonus payment instead, if the agent's effort was satisfactory. The results show that a non-binding, voluntary bonus payment leads to higher performance than an explicit incentive contract, which fines the agent for unsatisfactory performance.

However, performance provision does not only depend on monetary incentives but also seems to be influenced by situational factors such as appreciation, transparency, control, and closeness.

The positive effect of bonus payments on effort provision seems to be enhanced if the bonus payment is combined with an individual performance appraisal. A literature review of over

[4]Reciprocal behavior is also confirmed by further experimental (e.g., Fehr et al., 1993; Berg et al., 1995) and theoretical research (e.g., Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006).

300 papers done by Levy and Williams (2004) indicates that not only the monetary incentive but also the performance process itself increases job satisfaction and effort provision. The review shows that the supervisor's recognition and appreciation of the work performed are, aside from the monetary benefit, the main factor that causes an increase in an employee's job satisfaction. Kampkötter (2017) confirms the observation that the process itself plays an important role when it comes to the influence of bonus payments on performance by analyzing the relationship between performance appraisals and job satisfaction using data from the German Socio-Economic Panel (SOEP) study. Kampkötter finds that a monetary performance appraisal process conducted by the supervisor increases the employee's job satisfaction and leads to higher effort provision.

Ockenfels et al. (2015) show that transparency is playing a crucial part when it comes to bonus payments and job satisfaction. Perhaps surprisingly, Ockenfels et al. find that transparency does not increase job satisfaction per se, but can also amplify dissatisfaction. In a real-effort experiment, two agents received either a high or a low bonus payment for their performance. After the bonus payments were assigned, the agents played a Public Goods Game and a Dictator Game to transfer a part of their endowment to the principal in response to the bonus decision. Agents who received a higher bonus transferred 27.3% of their endowment while agents who received a lower bonus transferred 7.5% of their endowment. When the bonus payments were transparent (i.e. the agents got to know the percentage of the bonus budget they received), the response toward the principal was significantly more negative. In the treatment where the bonus payments were not transparent, the transfers did not differ significantly.

Experiments also indicate that employee effort is sensitive to the level of control. Fehr and Rockenbach (2003), and Fehr and List (2004) show that the principal's decision to use a punishment device leads to a decrease in effort provision by the agent in Trust Games. In a similar spirit, Falk and Kosfeld (2006), and Kajackaite and Werner (2015) find that controlling the agent has a negative influence on their performance. In a chosen-effort experiment conducted by Falk and Kosfeld, the agent had to choose a costly productivity activity that benefited the principal while the principal had the choice to either control (i.e. enforce a minimum effort) or trust the agent. The results show that the majority of the agents reduced their performance because most agents perceived control as a signal of distrust and low expectations by the principal. Kajackaite and Werner build upon the finding that control has a counterproductive effect on performance provision by showing that the principal's active decision to control affects the agent's kindness perception and triggers reciprocal responses. However, they find no significant change in the average output level in a real-effort experiment if the principal decides to implement a minimum performance requirement.

The effect of control on performance seems also be influenced by who is exercising control. Schmelz and Ziegelmeyer (2015) show that the effect of control depends on the closeness between the agent and the principal. By running an experimental principal-agent game in the internet as well as in a laboratory, Schmelz and Ziegelmeyer found that exercising control is less likely to reduce work performance in a remote than in a laboratory setting.

Control can not only be exerted by the principal, but also by a third party. Burdin et al. (2018) found agents to show higher effort provisions if principals abstain from control, than when a third party decides not to control. In an earlier work however, Burdin et al. (2015)

found quite the opposite: In an experiment based on the Principal-Agent Game used in Falk and Kosfeld (2006), the effort of the agent was higher if control was executed by a third party instead of directly by the principal.

The experimental finding that effort provision is not only influenced by monetary incentives but also by situational factors such as trust, transparency, and control is also supported by theoretical work. The model by Ellingsen and Johannesson (2005) shows that esteem influences performance, as a generous and trusting contract can elicit better performance from agents than a contract with low pay and strong incentives. Thereby, the model is based on the assumption that the principal's behavior conveys expectations about the agent and that these expectations might influence how the principal will rate the agent's performance ex-post. If the principal decides not to control the agent, the principal signals trust in the agent. This makes it harder for the agent to justify poor performance, in the sense that poor performance is not consistent with acceptable esteem. If the principal decides to control the agent, the principal shows a pessimistic expectation. In this case, the principal signals that low performance will not surprise them, which makes it easier for the agent to show low performance.

Previous research on interactions with machines, robots and automated processes suggests that these also have an impact on performance. By running five survey studies, Newman et al. (2020) show that reliance on algorithms is likely to have negative downsteam organizational consequences as people prefere humans over AI in HR, and argue that being evaluated by an automated process might evoke the percetion of reductionism, e.g. not taking certain qualitative factors or context into account. Gorny and Woodard (2020) found a negative and statistically significant correlation between automatability and job satisfaction. Corgnet et al. (2019) look at performance in a sequential task where participants had to work together with either another human or a robot – calibrated to the performance of an average worker – to fill out a grid. Human performance was significantly lower when the participants were matched with a robot compared to when matched with another human. The results show that humans perform better in a working environment with only humans than in a working environment in which they interact with a machine. To the best of my knowledge, there is no study that investigates the link between the use of automation and human work performance in an experimental setup.

# 3. Experimental design

Consider a modified two-stage Principal-Agent Game, similar to the design used in Falk and Kosfeld (2006), where the agent engages in a productive activity which is costly to the agent but beneficial to the principal. The agent has an initial endowment of 120 Points (1 Point equaled \$0.01 USD), while the principal's initial endowment is 0 Points. The agent chooses a productive activity $x$, and the cost of the productive activity for the agent is $c(x) = x$. The principal earns two times the agent's effort $p(x) = 2x$.[5] The principal then determines a threshold $x_t$ for a 'very good transfer' that the agent has to reach to get a bonus $b^*$. The

[5] An $p(x) = 2x$ mechanism is used as it allows the agent to form beliefs about the threshold set by the principal more easily than a complex mechanism.

bonus $b^* \in \{0, 120\}$ is paid by the experimenter. The agent receives the bonus if $x_t \geq x$. Thus, the payoff functions are $\prod_P = 2x$ for the principal and $\prod_A = 120 - x + b^*_{x_t}$ for the agent.

Automated decisions using algorithms, and decisions by humans essentially differ in one respect: algorithms are based on standardized processes and are built upon predetermined rules that are programmed *'ex-ante'* to the occurence while decisions by humans are more situation specific, hence, *'ex-post'* driven. To model the standardized process the following approach is used: if a non-automated - so called *manual Performance Appraisal System* (*manual PAS*) is used to determine the bonus payment, the principal knows the agent's productive activity $x$ before determining the minimal threshold $x_t$ for the agent to get the bonus $b^*$. Thus, the principle decides about the agent's performance threshold *'ex-post'*, after knowing the agent's actual performance. If an automated - so called *automated Performance Appraisal System* (*automated PAS*) is used, the principal does not know the agent's productive activity $x$ before determining the threshold $x_t$ for the agent to reach in order to get the bonus $b^*$. Therefore, the principle decides on the agent's performance threshold *'ex-ante'* to the performance, making the process less individualizable and more rigid.

## 3.1. Treatments

Depending on the treatment, either the principal (in treatment *HUMAN*) or a random mechanism (in treatment *SYSTEM*) decides whether to use the *manual PAS* or the *automated PAS* process.[6] In both treatments, the agents' efforts is elicited with the help of a strategy method.[7]

## 3.2. Procedure

The experiment was conducted online via Amazon Mechanical Turk (MTurk) using oTree (Chen et al., 2016). All sessions were run in August and September 2018 on Amazon MTurk using workers from the United States of America. The workers had to have completed at least 100 so-called Human Intelligence Tasks (HITs) on Amazon MTurk and had to have an approval rate of 99% for their completed HITs, to be able to take part in the experiment. All experimental stimuli and instructions were presented through a computer interface. Participants received a participation fee of $0.50 USD. A between-subjects design was used, so the data for all statistical tests are independent for the two treatments. The order in which both systems, the *manual PAS* and the *automated PAS*, were presented was randomly alternated for each participant in both treatments to control for potential order effects.

At the beginning of the experiment, all participants had to pass a test to ensure only humans would participate in the experiment. Therefore, the participants had to add up two two-digit numbers and write the correct answer into an input field. Participants who passed the human test were randomly assigned to a group of two as well as to a role. Each group

[6] In treatment *SYSTEM*, the a random mechanism decided to use either the *automated PAS* or the *manual PAS* with a probability of 50%.

[7] The results by Falk and Kosfeld (2006) do not indicate a difference between using the strategy method or the specific response method, and Charness et al. (2018) find qualitatively similar results for real-effort and stated-effort designs in a meta-study on effort measures in economic experiments. I therefore waived conducting an extra specific response treatment.

consisted of one agent (labeled participant A) and one principal (labeled participant B). Participants were first provided with the experimental instructions.[8] After reading the instructions, participants acted in four different stages: In stage 1, all participants had to answer a set of control questions to ensure they understood the instructions before proceeding. In stage 2, the productive activity, the task differed for principles and agents, also depending on the treatment. The agents had to decide how much of their initial endowment they wanted to transfer to their principals. In treatment *HUMAN*, the principals had to decide if they wanted to use the *manual PAS* or the *automated PAS*. In treatment *SYSTEM*, the random mechanism decided whether to use the *manual PAS* or the *automated PAS*. In stage 3, the bonus threshold task, each principal had to decide on the threshold for the amount that had to be transfered by the agent for him/her to get the bonus. In stage 4, the additional questions, the agents and the principals were asked to answer questions related to their expectations, their risk appetite and if they perceived the procedure to be fair. Participants were also asked about their age and gender. Furthermore, agents were asked about their thoughts on the transfer threshold set by the principals. As the accuracy of subjective beliefs is not of interest, agents' beliefs about the minimal transfer for getting a bonus were not incentivized.[9]

# 4. Behavioral predictions

| Social preference | Preferred productive activity ($x$) |
|---|---|
| Selfishness | 0 |
| Efficiency | 120 |
| Fair split | $\sim 60$ |
| Equality | $\sim 80$ |

The table shows possible social preferences and the corresponding preferred productive activities.

Table 1: Preferred productive activities ($x$) according to social preferences

Table 1 shows four social preferences, indicating that participants might prefer certain productive activities of the agent over others.

Assuming purely selfish preferences, the principal would not care whether the agent receives a bonus or not. Under this condition, agents maximize their payoff by choosing a transfer of $x = 0$. Assuming efficiency preferences the agent as well as the principal strive to maximize overall social welfare. The agent would transfer $x = 120$ points and the principal ensures a bonus payment of $b^* = 120$.

Assuming that the agent expects the probability for receiving a bonus to depend on the transfer, a more detailed analysis is required. The Model of Social Preferences by Charness and Rabin (2002) indicates that the amount transferred is influenced by social preferences and reciprocity. According to the model, people might prefere to have a higher monetary payoff

[8] Instructions can be found in the Appendix A.1.

[9] Not incentivizing beliefs also prevents hedging between the decision about how much to transfer and the expected payoff as a result of a correct belief.

than others or want to minimize the difference in payments between their own monetary payoffs and the payoff of others, and derive utility from reciprocal behavior. If the principal attaches importance to a fair split, the principal might demand an equal split of the agent's initial endowment and therefore prefers the agent to choose a productive activity of $x \approx 60$. On the other hand, allocative fairness models (e.g., Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) assume that an individual's utility is negatively influenced by an unequal outcome. Therefore, the principal would want the agent to choose a productive activity roughly around $x \approx 80$, ensuring an overall equal outcome. An utility-maximizing agent would therefore anticipate the preferences of the principal and choose a productive activitiy that matches the anticipated threshold set by the principal.[10]

Due to the fact that principals decide on the bonus threshold in both appraisal systems, the agents' expectations of the principal's behavior should be the same in both systems. I therefore expect no difference in the amount of points transferred by the agents between the *manual PAS* and the *automated PAS* within the two treatments (Hypothesis 1).

**Hypothesis 1** *The same amount of points is transferred in the manual PAS and in the automated PAS in*

*(i) treatment HUMAN, and in*

*(ii) treatment SYSTEM.*

While in treatment *SYSTEM* the system decides whether a *manual PAS* or an *automated PAS* is going to be used, the decision is incumbent on the principal in treatment *HUMAN*. More precisely, in treatment *HUMAN* the principal first decides about whether to use a *manual PAS* or an *automated PAS*, and then determines the threshold for the bonus payment. Due to this, in a manual PAS, the principal not only decides about a threshold but also determines immediately whether the agent receives a bonus or not. By actively choosing an *automated PAS*, the principal abstains from directly controlling whether a bonus is going to be paid or not, as the principal decides not to know the agent's productive activity before determining the threshold. In this respect, the principal's decision to use an *automated PAS* could be perceived as a lack of interest in or appreciation for the agent's productive activity.

Agents might also value the fact that the decision about whether the bonus is paid or not, is made by an individual knowing the agents actual productive activity. Thus agents might feel more appreciated in a *manual PAS* than in an *automated PAS*. From the theory on intention-based reciprocity by Rabin (1993), Dufwenberg and Kirchsteiger (2004), and Falk and Fischbacher (2006) we know that people tend to reward kind intentions and to punish

[10] The productive activity of the agent may also be infuenced by self- and social-image concerns, as well as risk preferences. From models of social image concerns (e.g., Bénabou and Tirole, 2006; Andreoni and Bernheim, 2009) and concepts of self-perception maintenance (e.g., Rabin, 1995; Beauvois and Joule, 1996) we know, that individuals perceive an unpleasant tension or disutility if their actions cause harm to their social-concept and/or self-concept of being a kind and fair individual. The agents anticipation of perceived disutility in not transferring anything might also affect the agent's productive activity. A risk-neutral agent is indifferent between all $x$. If the agent is risk-averse and believes that the probability of receiving a bonus is small, the agent tends to choose a smaller $x$.

unkind ones. Therefore, the agents might reciprocate by choosing a lower productive activity if the principal decides to use an *automated PAS*.

In treatment *SYSTEM*, a random mechanism instead of the principal decides whether to use an *automated PAS* or a *manual PAS*. Hence, the agent's reaction based on reciprocity should be reduced due to the lack of a direct counterpart who can be held accountable for the decision.

Based on the considerations above, agents are expected to choose a higher productive activity if the principal decides to use a *manual PAS* in treatment *HUMAN* compared to when a random mechanism chooses the *manual PAS* in treatment *SYSTEM*. Correspondingly, agents are expected to choose a lower productive activity if the principal decides to use an *automated PAS* in treatment *HUMAN* compared to when a random mechanism chooses the *automated PAS* in treatment *SYSTEM* (Hypothesis 2).

**Hypothesis 2** *In treatment HUMAN,*

*(i) more points are transferred in a manual PAS, and*

*(ii) fewer points are transferred in an automated PAS*

*than in treatment SYSTEM.*

# 5. Results

Overall, 520 participants (44.4% female) contributed to the study.[11] The participants were on average 37 years old. The study took about 10 minutes to complete and the participants earned on average $1.8 USD.

As a between-subjects design for the treatments was used, the data for all statistical tests are independent for the different treatments. For the different *PAS*s a within-subjects design is used, i.e. agents were asked about their transfer in a *manual PAS* as well as in an *automated PAS* using a strategy method.

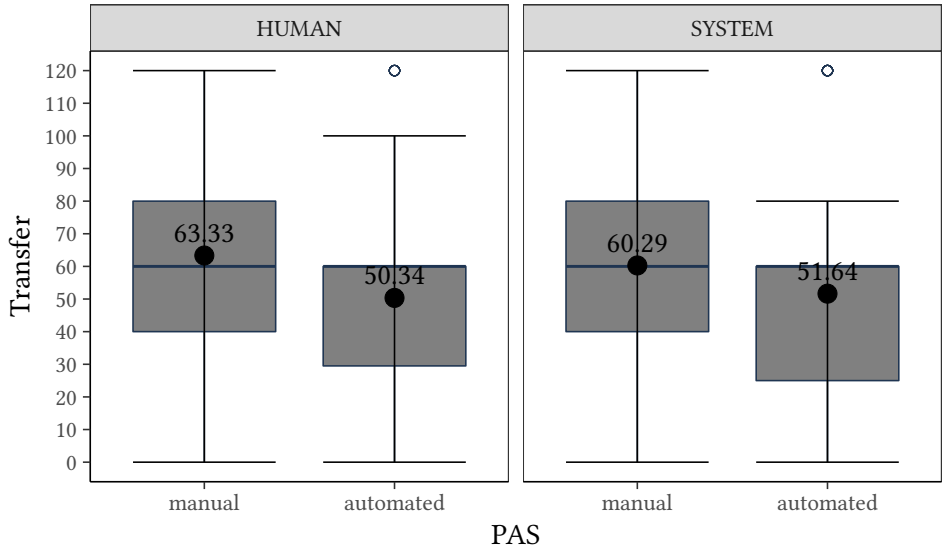
In the following section, the number of points transferred by the agent within each treatment is analyzed followed by a comparison of the transferred points in the *manual PAS* and the *automated PAS* between treatments.[12]

## 5.1. Hypothesis 1: *manual PAS* vs. *automated PAS*

According to Hypothesis 1, the agents should transfer the same amount of points to the principals under a *manual PAS* and under an *automated PAS* in both treatments.

[11] In treatment *HUMAN* 259 participants (42% female) (131 agents and 128 principals) and in treatment *SYSTEM* 261 participants (46.7% female) (135 agents and 126 principals) took part in the experiment. The number of agents differs from the number of principals as some principals left the experiment before setting a threshold. In this case, the experimenter granted the bonus to the remaining agents independent of their productive activity.

[12] An analysis of the principals' behavior can be found in Appendix A.4.




Filled dots represent means, lines represent medians.

Figure 1: Box-and-whisker plots for the transferred points.

| | *HUMAN* | *SYSTEM* |
|---|---|---|
| *manual PAS - automated PAS* | $\Delta = 12.99$ | $\Delta = 8.65$ |
| | ($p$ =0.0000) | ($p$ =0.0000) |

The table shows differences between the *PAS*s ($\Delta = \ldots$) and $p$-values for a two-sided Wilcoxon signed-rank test. The tests report no problem about the frequency of ties.

Table 2: Differences in the agents' transferred points between *PAS*s.

As Figure 1 shows, the agents transfered on average more to the principals in the *manual PAS* than in the *automated PAS* in both treatments.[13]

Table 2 shows the difference in the mean number of transferred points between the *manual PAS* and the *automated PAS*, and provides the corresponding $p$-values for whether the means differ significantly within the treatments. The table confirms that the agents transfer significantly more to the principal in the *manual PAS* than in the *automated PAS* in both treatments.

Hence, Hypothesis 1.(i) and Hypothesis 1.(ii), i.e. that the same amount of points are transferred in a *manual PAS* and in an *automated PAS* for treatment *HUMAN* as well as for treatment *SYSTEM*, can not be confirmed.

[13] An analysis of the frequency of the agents who choose fewer, more or the same in an *automated PAS* than in a *manual PAS* can be found in Table 6 in Appendix A.3.

## 5.2. Hypothesis 2: treatment *HUMAN* vs. treatment *SYSTEM*

According to Hypothesis 2.(i), the agents should transfer more points in the *manual PAS* and, according to Hypothesis 2.(ii), the agents should transfer fewer points in the *automated PAS* in treatment *HUMAN* than in treatment *SYSTEM*. In fact, this is what we see in Figure 1.

Table 3 provides the difference in the mean number of transferred points for both *PAS*s and $p$-values for whether the means differ significantly between the treatments. Indeed, agents in treatment *HUMAN* transfer on average more points in the *manual PAS* and fewer points in the *automated PAS* than agents in treatment *SYSTEM*. However, the difference is not statistically significant in either the *manual PAS* or the *automated PAS*. Hence, Hypothesis 2.(i) or Hypothesis 2.(ii), can not be confirmed.

| | *HUMAN - SYSTEM* |
|---|---|
| *manual PAS* | $\Delta$ = 3.04 |
| | ($p$ = 0.1358) |
| *automated PAS* | $\Delta$ = -1.3 |
| | ($p$ =0.4819) |

The table shows differences between the *PAS*s ($\Delta$ = . . .) and $p$-values for a one-sided Wilcoxon rank-sum test. The tests report no problem about the frequency of ties.

Table 3: Differences in the transferred points in *manual PAS* and automated PAS between the treatments.

## 5.3. Agents' expectations

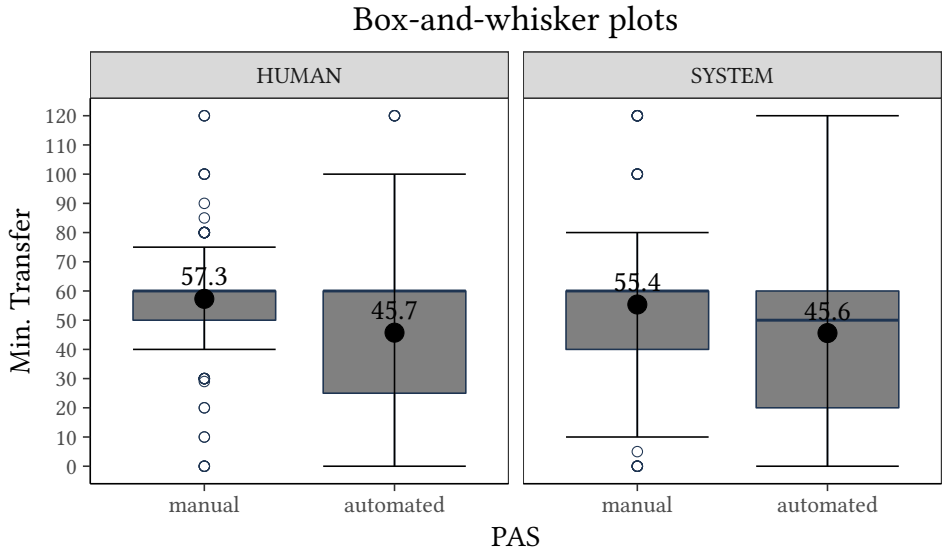
Agents were asked about their expectations for getting a bonus. As Table 4 shows, most agents expected to get a bonus but agents in treatment *SYSTEM* tended to be overall less optimistic about receiving a bonus than agents in treatment *HUMAN*.

Furthermore, as Figure 2 shows and Table 5 confirms, agents in both treatments expected the threshold to be significantly higher in the *manual PAS* than in the *automated PAS*.

| | HUMAN | SYSTEM |
|---|---|---|
| Strongly agree | 9.20 | 17.20 |
| Agree | 77.30 | 65.60 |
| Disagree | 10.10 | 15.60 |
| Strongly disagree | 3.40 | 1.60 |

See Question 5 from Appendix A.2.

Table 4: Agents' beliefs about receiving a bonus [%].




Filled dots represent means, lines represent medians.
See Question 3 and 4 from Appendix A.2.

Figure 2: Box-and-whisker plots for agents' beliefs about the threshold set by the principal.

| | *HUMAN* | *SYSTEM* |
|---|---|---|
| *manual PAS - automated PAS* | $\Delta = 11.57$ | $\Delta = 9.78$ |
| | ($p$ =0.0000) | ($p$ =0.0001) |

The table shows differences between the *PAS*s ($\Delta = \ldots$) and $p$-values for a two-sided Wilcoxon signed-rank test if this difference could be zero.

Table 5: Differences in the agents expectations about the threshold between *PAS*s.

# 6. Conclusion

To what extent does automating management decisions present companies with new challenges e.g.,what are the hidden costs of automation? I found that performance is significantly lower under an *automated PAS* than under a *manual PAS*. However, performance was not influenced by whether a person or a system decided to use a specific *PAS*.

The presented experiment studies whether automation leads to a decrease in employees' performance. For this purpose, I set up a Principal-Agent experiment in a real job market (Amazon MTurk) and compared performance under an *automated PAS* and a *manual PAS*. Furthermore, I investigated whether it matters if the company (treatment *SYSTEM*) or the direct supervisor (treatment *HUMAN*) decides on what *PAS* to use.

I find that the agents' performance is significantly lower under an *automated PAS* than under a *manual PAS* in both treatments. Thus, I observe hidden costs of automation in the form of lower performance when using automation (*automated PAS*) compared to human-to-human interactions (*manual PAS*).

Based on standard theoretical models, one can not expect a difference between the performance in both PASs. A possible explanation for this can be found in different expectations of the agents regarding the performance threshold set by the principles in both PAS. The results show that the agents expect to have to perform higher when being evaluated ex-post by a human compared to an ex-ante defined algorithm. Hence, agents might adjust to their higher expectations by showing higher effort provisions.[14]

Under the assumption that employees care about who decides on using automation rather than a manual process, we expect the performance to differ depending on whether the principal (e.g. supervisor) or system (e.g. company) decided to use an *automated PAS*. I find, however, no significant difference in the performance if the principal or system makes the decision. Hence, the hidden costs of automation seem to be independent of who decides to automate.

As the difference between a human decision and a decision made by a system is quite subtle in an online experiment (as all interaction takes place via a computer interface), the divergence between treatments might not have been salient enough to the participants, explaining the similarity in the performance in both treatments. Further research would be needed to prove this claim.

The results show that, besides the tremendous benefits automation generates, it might also have some downsides. Superiors in charge of implementing an automated system which replaces human-to-human interactions may benefit from communicating that the underlying parameters and demands of the automated system do not differ from the demands set by humans. In addition, developers and other responsible decision makers should not only consider the benefits, such as time and cost reduction, when implementing automated systems, but also be aware of the hidden costs of automation, such as negative impacts on motivation and performance of employees and others interacting with such algorithms, software and/or AI.

In conclusion, the results show that hidden costs exist, that should be considered when replacing human-to-human interactions with automated processes.

# References

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Beauvois, J.-L. and Joule, R. (1996). *A radical dissonance theory*. Taylor & Francis, London, GB.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

[14]However, as I analyze in the Appendix, principals set more or less the same thresholds in a *manual PAS* and in an *automated PAS*. This discrepancy between the agents' expectations and the principals' actual behavior leaves some room for further exploration.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193.

Burdin, G., Halliday, S., and Landini, F. (2015). Third-party vs. second-party control: Disentangling the role of autonomy and reciprocity.

Burdin, G., Halliday, S., and Landini, F. (2018). The hidden benefits of abstaining from control. *Journal of Economic Behavior & Organization*, 147:1–12.

Chakraborti, T., Isahagian, V., Khalaf, R., Khazaeni, Y., Muthusamy, V., Rizk, Y., and Unuvar, M. (2020). From robotic process automation to intelligent process automation. In *International Conference on Business Process Management*, pages 215–228. Springer.

Charlier, R. and Kloppenburg, S. (2017). Artificial intelligence in hr: a no-brainer. Retrieved December 30, 2018, from `https://www.pwc.nl/nl/assets/documents/artificial-intelligence-in-hr-a-no-brainer.pdf`.

Charness, G., Gneezy, U., and Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, 149(3):74–87.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otree — an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Corgnet, B., Hernán-Gonzalez, R., and Mateo, R. (2019). Rac(g)e against the machine? social incentives when humans meet robots. *GATE Working Paper No. 5824*.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Ellingsen, T. and Johannesson, M. (2005). Trust as an incentive. *Stockholm School of Economics Working Paper mimeo*.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459.

Fehr, E., Klein, A., and Schmidt, K. M. (2007). Fairness and contract design. *Econometrica*, 75(1):121–154.

Fehr, E. and List, J. A. (2004). The hidden costs and returns of incentives — trust and trustworthiness among ceos. *Journal of the European Economic Association*, 2(5):743–771.

Fehr, E. and Rockenbach, B. (2003). Detrimental effects of sanctions on human altruism. *Nature*, 422(6928):137–140.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schmidt, K. M. (2007). Adding a stick to the carrot? the interaction of bonuses and fines. *American Economic Review*, 97(2):177–181.

Gorny, P. M. and Woodard, R. C. (2020). Don't fear the robots: Automatability and job satisfaction. *MPRA Paper, No. 103424*.

Kajackaite, A. and Werner, P. (2015). The incentive effects of performance requirements – a real effort experiment. *Journal of Economic Psychology*, 49:84–94.

Kampkötter, P. (2017). Performance appraisals and job satisfaction. *The International Journal of Human Resource Management*, 28(5):750–774.

Kaur, N. and Sood, S. K. (2017). A game theoretic approach for an iot-based automated employee performance evaluation. *IEEE Systems Journal*, 11(3):1385–1394.

Levy, P. E. and Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, 30(6):881–905.

Manyika, J., Chui, M., Miremadi, M., and Bughin, J. (2017). A future that works: Ai, automation, employment, and productivity. *McKinsey Global Institute Research, Tech. Rep*, 60:1–135.

Markets and Markets (2020). Business process management market by component, deployment type, organization size, business function (sales and marketing, hrm, procurement and scm, and customer service support), industry, and region - global forecast to 2025. Retrieved December 30, 2021, from `https://www.marketsandmarkets.com/Market-Reports/business-process-management-market-157890056.html`.

Mohanty, S. and Vyas, S. (2018). Intelligent process automation= rpa + ai. In *How to Compete in the Age of Artificial Intelligence*, pages 125–141. Springer.

Newman, D. T., Fast, N. J., and Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, 160:149–167.

Ockenfels, A., Sliwka, D., and Werner, P. (2015). Bonus payments and reference point violations. *Management Science*, 61(7):1496–1513.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. *Department of Economics UCB Working Paper No. 95-241*.

Schmelz, K. and Ziegelmeyer, A. (2015). Social distance and control aversion: Evidence from the internet and the laboratory. *Thurgau Institute of Economics and Department of Economics at the University of Konstanz Research Paper Series, No.100*.

Stone, H. S. (1971). *Introduction to computer organization and data structures*. McGraw-Hill, New York, NY.

Upadhyay, A. K. and Khandelwal, K. (2018). Applying artificial intelligence: implications for recruitment. *Strategic HR Review*.

Yen, I.-L., Bastani, F., Huang, Y., Zhang, Y., and Yao, X. (2017). Saas for automated job performance appraisals using service technologies and big data analytics. In *2017 IEEE International Conference on Web Services (ICWS)*, pages 412–419. IEEE.

## Statements and Declaration

This work was funded by the Max Planck Society through the International Max Planck Research School on Adapting Behavior in a Fundamentally Uncertain World. The author has no relevant financial or non-financial interests to disclose.

# A. Appendix

This section contains additional information on the interfaces and questions used in the treatments. I also present further analyses of the data I collected in addition to the data used to test my hypotheses. Data and methods are available online.

## A.1. Instructions

*Note: The following instructions are for treatment HUMAN. Differences for treatment SYSTEM are added in italic.*

This HIT [Human Intelligence Task] is an economic experiment. Please read the following instructions carefully. The instructions provide you with all the information required for participating in the experiment. You will receive $0.50 USD for participating in the experiment (paid only if you finish the experiment). Your final payoff is the $0.50 USD for participating in the experiment plus the amount earned during the experiment. You will earn at least the $0.50 USD for participating in the experiment. In the experiment, the currency used is points. Your points will be converted to USD at the end of the experiment using a conversion rate of **1 point = $0.01 USD**.

**General setup**

In this experiment, you are matched with another human participant. You will play in a group of two. All decisions are made anonymously. No participant knows with whom (s)he is matched. During the experiment, the members of the group are called ”participant A” and ”participant B”. The roles are randomly assigned.

**The experiment**

Participant A starts with 120 points at the beginning of the experiment. Participant B starts with no points. Each participant has to make a decision during the experiment. The decisions are explained below. Please read the explanations for both participants as both decisions will affect the number of points you will earn.

**Participant A's decision:**

Participant A has to decide how many points (s)he wants to transfer to participant B. The points transferred to participant B are doubled by the experimenter, meaning each point transferred to participant B reduces the points of participant A by one point but increases the points of participant B by two points.

**Participant B's decision:**

Before participant B knows what participant A transferred, participant B *[the system]* selects an approach. The two possible approaches (Approach BLUE or Approach GREEN) are explained below.

In each approach participant B has to decide if participant A should be given a **bonus of 120 points** for a **”very good transfer”**. The bonus is paid by the experimenter and does not reduce the points of participant B. The difference between the two approaches is how participant B determines the minimum amount (threshold) that participant A has to transfer

to get a bonus.

In Approach BLUE, participant B **knows** the amount transferred by participant A when determining the threshold. The decision screen will look like this:

| |
|---|
| You *[the system]* decided to use Approach BLUE.<br><br>**Participant A has transferred X of 120 points to you.**<br><br>If participant A has transferred at least the threshold amount (s)he gets a bonus of 120 points (paid by the experimenter).<br>Please indicate your threshold here:<br><br>……Points |

In Approach GREEN, participant B **DOES NOT know** the amount transferred by participant A when determining the threshold. The decision screen will look like this:

| |
|---|
| You *[the system]* decided to use Approach GREEN.<br><br>If participant A has transferred at least the threshold amount (s)he gets a bonus of 120 points (paid by the experimenter).<br>Please indicate your threshold here:<br><br>……Points |

**Further note:**
Participant A has different fields to enter amounts in case Approach BLUE or Approach GREEN is used.
***Some examples:***

- ***Example 1:*** Participant A transfers 0 points to participant B. Participant A will have 120 points (120 - 0) plus eventually a bonus of 120 points. Participant B will have 0 points (0 x 2). In addition, both participants receive $0.50 USD for participating.
- ***Example 2:*** Participant A transfers 40 points to participant B. Participant A will have 80 points (120 - 40) plus eventually a bonus of 120 points. Participant B will have 80 points (40 x 2). In addition, both participants receive $0.50 USD for participating.
- ***Example 3:*** Participant A transfers 80 points to participant B. Participant A will have 40 points (120 - 80) plus eventually a bonus of 120 points. Participant B will have 160 points (80 x 2). In addition, both participants receive $0.50 USD for participating.

- ***Example 4:*** Participant A transfers 120 points to participant B. Participant A will have 0 points (120 - 120) plus eventually a bonus of 120 points. Participant B will have 240 points (120 x 2). In addition, both participants receive $0.50 USD for participating.

Before clicking "Next" please make sure you have read and understood the instructions. After clicking "Next" we will match you with the next person starting the experiment. This might take some time.

## A.2. Questions

*Note: All participants were asked to complete a questionnaire. The questions were asked right after the decision and before the final outcome was announced. The answer method used is presented in brackets. Apart from the first four questions, which were only presented to agents, all questions were asked to agents and principals.*

1. Why did you choose to transfer the amount you have chosen to participant B in Approach BLUE (participant B **knows** how much you transferred)? [Open Question] *(For the answers given see online data-set)*

2. Why did you choose to transfer the amount you have chosen to participant B in Approach GREEN (participant B **does not know** how much you transferred)? [Open Question] *(For the answers given see online data-set)*

3. What do you think is the minimum amount you would have had to transfer to get the bonus if participant B decided to use Approach BLUE (participant B **knows** how much you transferred)? [Integer from 0 to 120 points] *(For an analysis of the answers given see Section 5.1)*

4. What do you think is the minimum amount you would have had to transfer to get the bonus if participant B decided to use Approach GREEN (participant B **does not know** how much you transferred)? [Integer from 0 to 120 points] *(For an analysis of the answers given see Section 5.1)*

5. How much do you agree with this statement: 'I think that I will get a bonus.'? [*"Strongly disagree"; "Disagree", "Agree"; "Strongly agree"*] *(For an analysis of the answers given see Appendix 5.3)*

6. Do you consider the procedure to get the bonus to be fair? [*"YES"; "NO"*] *(For an analysis of the answers given see Appendix A.5)*

7. How do you see yourself: Are you a person who is willing to take risks or do you try to avoid taking risks? Please select a number on a scale from 0 to 10. The value 0 means: 'not at all willing to take risks' and the value 10 means: 'very willing to take risks'. [scale 0 to 10] *(For an analysis of the answers given see Appendix A.6)*

8. Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with other people? Please select a number on a scale from 0 to 10. The value 0 means: 'can't be too careful' and the value 10 means: 'most people can be trusted'. [Scale 0 to 10] *(For an analysis of the answers given see Appendix A.6)*

9. What is your gender? [*"MALE"; "FEMALE"; "OTHERS"*] *(For an analysis of the answers given see Section 5)*

10. What is your age [in years]? [*Integer*] *(For an analysis of the answers given see Section 5)*

## A.3. Relative frequency of the agents' transfer decision

| | HUMAN | SYSTEM |
|---|---|---|
| less | 42.70 | 30.40 |
| more | 6.90 | 6.70 |
| same | 50.40 | 63.00 |

The table shows the percentage of agents transferring the same, more, or less in an *automated PAS* than in a *manual PAS* by treatment.

Table 6: Agents' transfer decisions [%].

Table 6 shows the relative frequency of agents who transferred less, more, or the same in an *automated PAS* than in a *manual PAS*. The table reveals that around half of the agents transferred the same in a *manual PAS* as in an *automated PAS* in both treatments. Nevertheless, around 40% of the agents transferred fewer points in an *automated PAS* than in a *manual PAS* in treatment *HUMAN* and slightly less than one-third of the agents did so in treatment *SYSTEM*.

## A.4. Analysis of the principals' behavior

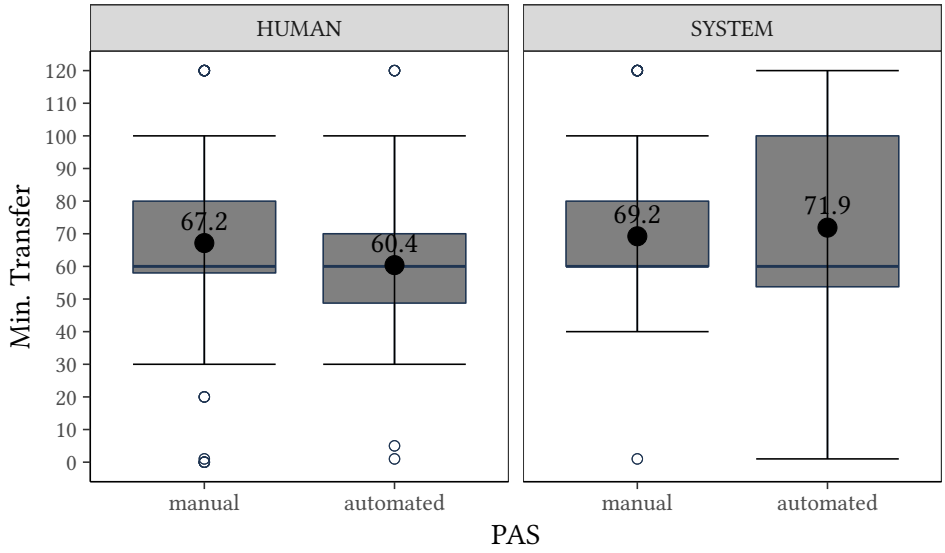
As Table 7 shows, the vast majority of the principals decided to use a *manual PAS* instead of an *automated PAS* in treatment *HUMAN*, where the principals were able to choose.

| | HUMAN | SYSTEM |
|---|---|---|
| manual | 78.10 | 49.20 |
| automated | 21.90 | 50.80 |

The table shows the percentage of principals choosing a *manual PAS* or *automated PAS*.

Table 7: *PAS* choices by principals and the system[%].

As suggested by Figure 3 and confirmed by Table 8, the threshold set by the principals in the *automated PAS* does not differ significantly from the threshold set in the *manual PAS* in both treatments.




Filled dots represent means, lines represent medians.

Figure 3: Box-and-whisker plots for thresholds set by principals.

| | *HUMAN* | *SYSTEM* |
|---|---|---|
| *manual PAS - automated PAS* | $\Delta$ = 6.7571429 | $\Delta$ = -2.6491935 |
| | ($p$ =0.1953) | ($p$ =0.8511) |

The table shows differences between the *PAS*s ($\Delta$ = ...) and $p$-values for a two-sided Wilcoxon rank-sum test where this difference could be zero.

Table 8: Differences in the principals' threshold set between *PAS*s.

| | *HUMAN - SYSTEM* |
|---|---|
| *manual PAS* | $\Delta$ = -2.0758065 |
| | ($p$ = 0.9565) |
| *automated PAS* | $\Delta$ = -11.4821429 |
| | ($p$ =0.2283) |

The table shows differences between the *PAS*s ($\Delta$ = ...) and $p$-values for a two-sided Wilcoxon rank-sum test.

Table 9: Differences in the principals' threshold set in *manual PAS* and *automated PAS* between the treatments.

As Table 9 shows, the threshold does not differ significantly between both treatments. The thresholds set in *automated PAS*s in treatment *SYSTEM*, however, are more dispersed than in the other conditions. In summary, principals in treatment *HUMAN*, who decided on their own which approach to use, did not set a significantly different threshold than participants in treatment *SYSTEM*, where the system decided randomly which approach to use.

## A.5. Perceived fairness of the procedure

I asked the participants if they perceive the procedure to result in a fair outcome (see Question 6). In treatment *HUMAN*, 84.87% of the participants perceived the procedure to be fair. In treatment *SYSTEM*, the procedure was perceived to be fair by 80.33% of the participants. As Table 10 shows, the assessment by the agents and the principals hardly differs.

| | Agent | Principal |
|---|---|---|
| *HUMAN* | 84.87 | 83.59 |
| *SYSTEM* | 80.33 | 83.33 |

Table 10: Assessment of the fairness of the procedure [%].

## A.6. Participants' propensity for risk and trust

All participants were asked if they are a person who is willing to take risks or tries to avoid taking risks (see Question 7) and if they would say that most people can be trusted or that you cannot be too careful in dealing with other people (see Question 8). Willingness to take risks was measured by a continuous scale from 'not at all willing to take risks' (0) to 'very willing to take risks' (10). The level of trust was measured by a continuous scale from 'can't be too careful' (0) to 'most people can be trusted' (10). As Table 11 shows, agents and principals were slightly risk averse and somewhat concerned about the trustworthiness of other people.

| | Agent | Principal |
|---|---|---|
| Risk | $\varnothing = 4.36$ | $\varnothing = 4.53$ |
| | (2.58) | (2.3) |
| Trust | $\varnothing = 4.9$ | $\varnothing = 5.04$ |
| | (2.55) | (2.41) |

The table shows the means for risk and trust ($\varnothing = \ldots$) and the corresponding standard deviations (in brackets).

Table 11: Mean and standard deviation for levels of risk and trust.