**3. METHOD**

We conducted a between-subjects online experiment in July and August of 2017. Participants examined a managerial scenario, (Mintzberg, 1975) in which a decision was made by either humans, algorithms, or a combination of both. We examined how the type of decision-maker would influence people’s perceptions of the decisions by collecting both quantitative ratings of the decisions and qualitative reasons behind those ratings. We used a scenario-based method, which is commonly used in social psychology and ethics research, to investigate people’s opinions, beliefs, and attitudes (e.g., Petrinovich, O’Neill, and Jorgensen, 1993); studies have suggested consistency between people’s behaviors in scenario-based experiments and their behaviors in real life (Woods et al., 2006).

**3.1 Participants**

We recruited participants on Amazon Mechanical Turk (MTurk) to take an online survey that took 6.4 minutes on average to complete. Participants had to reside in the US, be at least 18 years old, have completed at least 100 HITs (Human Intelligence Task, MTurk’s task unit), and have at least a 95% HIT approval rate. Participants were compensated $1.00 for their time, a rate more than the minimum wage in the US ($7.25/hour). 1637 people responded. We omitted participants, indicated that they did not reside in the US (N=7), were younger than 18 (N=118), did not pass the first attention check (N=153), or did not finish the survey (N=7). These procedures left 1352 participants in the sample, about four-fifths of the original number. From these 1352 participants, we omitted those that failed the reading check (N=18) and failed the manipulation check (N=88). However, we allowed participants that failed the manipulation check but had correct explanations of the situation (N=18). We also omitted participants that correctly answered the manipulation check but failed to correctly explain the situation (N=90). In the end, there were 1174 participants in the sample.

Their ages ranged from 18 to 88, with an average age of 35.9 (SD = 11.39). 51% of the sample was female. They were fairly well educated: the mean education score was 4.12 (SD = 1.33; 3=”Some college or currently enrolled”, 4=”2-year degree”). The sample was 76% Caucasian, 8.28% Asian, 6.06% Hispanic, 7.42% African American, 0.43% Native American, 0.26% Pacific Islander, and 1.37% other.

**3.2 Materials**

We presented participants with four scenarios in which either algorithmic or human managers or a combination of both made decisions that had significant impacts on human workers (Table 1). Our scenarios were based on real-world situations in which algorithms are being employed in middle-management roles linked to the main functions of management (Mintzberg, 1975). We organized these four tasks based on the amount of “mechanical” skills and the amount of “human” skills involved, with assignment being the most mechanical and evaluation requiring the most human skills. The work assignment scenario included maintenance tasks in a factory, which used algorithms to predict the likeness of machinery components malfunctioning in order to prevent a larger stoppage in the overall workflow (Jardine, Lin, and Banjevic, 2006) (Table 1.a). The scheduling scenario used the example of an employee shift scheduling algorithm, which determined when café baristas would be called into work based on the predicted number of customers in the café at a given time and the baristas’ schedule preferences (Table 1.b) (Kantor, 2014; Pinedo, Zacharias, and Zhu, 2015).

**Table 1. Managerial scenarios presented to participants**

The hiring scenario was based on job search websites, such as LinkedIn, that use algorithms to analyze resumes and select top candidates for onsite interviews (Table 1.c). Finally, the work evaluation scenario revolved around a customer service call center that used a natural language-based algorithm to evaluate the performance of its employees (Petrushin, 1999) (Table 1.d).

We used the projective, third-person viewpoint while creating scenarios so that participants were reading scenarios that describe another person’s experience (e.g., Chris works […]) as supposed to ones that directly put the readers into the scenario (e.g., you work […]). The projective viewpoint has been shown to minimize social desirability effects, which is the desire to present socially desirable answers rather than honest opinions, and to have considerable external validity (Nisbett, et al., 1973). The scenarios had two parts. The first part described the general policy of management, and the second part described a specific instance of the policy involving a worker. We manipulated the type of decision-maker (algorithmic, human, or both) for all scenarios.

We conducted a pilot test to check which decision-makers people thought would perform each task better and whether those perceptions were in line with our assumptions; specifically, whether people thought the combination of the algorithm and the human would outperform the algorithm and/or the human alone. The within-subjects survey (N=21) presented the general policy portion of the four scenarios in a random order and asked which type of decision-maker would perform the task better (1=“Algorithms will perform the task better than human managers,” 2= “Both algorithms and human managers will perform the task equally well,” 3= “Human managers will perform the task better than algorithms”). We ran a multi-level analysis controlling for individuals for repeated measures. The results show that people thought that algorithms would perform the work assignment and scheduling tasks slightly better the human managers (Massign=1.4 (SE=.15), Mschedule=1.65 (SE=.15)), and that human managers would perform the hiring and evaluation tasks better than algorithms (Mhire=2.5 (SE=1.5), Mevaluate=2.9 (SE=1.5)).

**3.3 Procedure**

After receiving consent and affirming that the participants were over the age of 18 and a US resident, we gave the participants an attention check question adopted from Egelman and Peer, 2015. Failure to correctly answer the attention check resulted in immediate disqualification from the survey. Those who passed the check were randomly assigned to a decision-maker [human/algorithmic/both] in one of the four task scenarios. Participants in the algorithmic condition were shown this definition of “algorithm”: “Algorithms are processes or sets of rules that a computer follows in calculations or other problem-solving operations. In the situation below, an algorithm makes a decision autonomously without human intervention.”[[1]](#footnote-1) We presented the definition to ensure that participants understood our usage of the word algorithm, and bolded the analyst and the decision-maker of each scenario in order to emphasize any differences or similarities. Participants were then presented with a scenario and survey questions about their thoughts on the scenario they had just read. The manipulation check and demographic questions were asked at the end.

**3.4 Measures**

**3.4.1 Perceptions of decisions**

Except for a few open-response questions, all survey items used a 7-point Likert-type scale. Response options varied based on the questions (i.e., Strongly Disagree, Very Unfair).

**Decision fairness.** The question on the decision’s fairness was adopted from previous research (Brockner et al., 1994; Konovsky and Folger, 1991): “How fair or unfair is it for [scenario subject] that the [algorithm/manager takes the action specified in the scenario]?” For example: “How fair or unfair is it for Chris that the manager evaluates his performance? The scale ranged from “Very unfair” (1) to “Very fair” (7).

**Trust.** To ascertain subjects’ trust in the reliability and accuracy of the decision presented in each scenario, we asked, “How much do you trust that the [algorithm/manager] make good-quality [decision specified in the scenario]?” The scale ranged from “No trust at all” (1) to “Extreme trust” (7).

**Emotional response.** To understand how participants thought the decision affected the scenario’s subject, we asked how much they agreed or disagreed that the decision-maker’s decision would make the scenario’s subject feel happy, joyful, proud, disappointed, angry, and frustrated (1=“Strongly disagree”, 7=“Strongly agree”) (Larsson, 1987; Weiss, Suckow, and Cropanzano, 1999). We constructed an emotional response scale by averaging answers to the three positive adjectives and the reversed answers to the negative adjectives, so that greater numbers meant more positive emotion. The scale was very reliable (Cronbach’s α = .9).

**Open-response question.** After each of the three questions above, we asked participants to explain their reasons for their numeric ratings.

**3.4.2 Manipulation checks, attention checks, and demographic questions**

At the end of the survey in the algorithmic decision conditions, participants were asked an open-ended question: “In your own words, please briefly explain what you think algorithms are.” The answers confirmed that participants perceived algorithms as autonomous decision-makers. Another manipulation check question asked all participants: “Which of the following made the decisions in the situations that you read?” and provided a choice between humans, algorithms, or both. All participants correctly answered this question by condition.

We used one attention check (Egelman and Peer, 2015) in the beginning of the survey to immediately disqualify participants. Throughout the survey, we also asked participants how much they agreed with the statement, “I do not read the questions in this survey.” At the end of the survey, we also asked participants to indicate their knowledge of algorithms (1=”No knowledge at all”, 5=“Expert knowledge of algorithms”) along with the demographic questions.

**3.5 Analysis**

We conducted a one-way ANOVA to find the main effect of the decision-maker on perception of the decision for each decision scenario and a multi-level analysis on the main and interaction effects of the decision-maker type and task types. We qualitatively analyzed participants’ reasons for their answers to the three questions about fairness, trust, and emotional response (Strauss and Corbin, 1990). We open-coded data at the response level in conjunction with participants’ survey ratings to identify emerging themes. We grouped different themes to explain how participants responded to and judged the decisions depending on the type of decision-maker.

1. Oxford Dictionaries http://www.oxforddictionaries.com/ [↑](#footnote-ref-1)