Operational Transparency Under System Failure

Jorge Mejia
Kelley School of Business, Indiana University, Bloomington, IN 47405, jmmejia@iu.edu, http://go.iu.edu/jorge

Chris Parker
Kogod School of Business, American University, Washington, DC 20016, chris.parker@american.edu, https://www.chrisparker.io

---

Many organizations are starting to realize the value of optimizing the collaboration between humans and AI systems (Wilson and Daugherty 2018). In service operations, there is an increasing dependence on artificial intelligence (AI) systems—a trend that is likely to continue in the coming years (Olsen and Tomlin 2019). At their core, these services rely on vast amounts of training data that is often curated by humans. . However, like most computer systems, the systems used to create the training data are not perfect; they can crash, lag, and be hampered by implementation problems of typical software projects. Our primary research question is to what extent AI system failures can impact worker performance and ultimately affect the successful collaboration of humans and machines.

If indeed there are reductions in worker performance when systems fail, what remedies may work for managers? Two possible options from the literature are 1) operational transparency, and 2) performance-based pay. Providing operational transparency has been shown to improve consumers' perceived quality (Buell and Norton 2011). In addition, Buell et al. (2017) examine the impact of communicating successful outcomes to customers. They find that people are more likely to engage with the government in the future when their problems are marked as resolved. Finally, it would be relatively cheap to implement in the interface between humans and AI systems, but does it help improve worker performance after a system failure? Alternatively, performance-based pay is an economic lever that increases payment to workers based on the quantity or quality of their output. Previous research has shown that performance-based pay can increase worker quality in digital work settings (Ho et al. 2015). However, can financial incentives to provide high-quality work fully mitigate the negative consequences of a system failure? Or more generally, how do operational transparency and financial incentives compare as tools to improve the human-AI interface?

Table 1 Overview of Experiments

Experiment Number	Classification Task	Incentive Scheme	Subjects	Timing
1	Images	Single Pay	Students (In Person)	Nov. 2018
2	Images	Single Pay	mTurk	Nov. 2018
3	Text	Single Pay	Students (In Person)	Nov. 2019
4	Text	Single Pay	mTurk	Nov. 2019
5	Images	Threshold Pay	Students (Online)	Dec. 2019
6	Images	Threshold Pay	mTurk	Dec. 2019
7	Text	Threshold Pay	Students (Online)	Dec. 2019
8	Text	Threshold Pay	mTurk	Dec. 2019

Overview of the experiments performed. In all experiments, subjects were asked demographic questions and questions aimed at measuring self confidence (PEI and TROSCI), proverbs, effort, satisfaction, and cognitive ability. In addition, subjects played a game following experiments 3-8.

To explore the impact of system failure on digital worker quality and the potential for operational transparency to serve as a solution to any negative effects, we conduct multiple experiments where subjects are asked to perform the most popular AI task: image classification. Specifically, subjects were repeatedly shown pictures of flowers and asked to classify them as daisy, dandelion, rose, sunflower, or tulip varieties using the Google AI image classification toolkit.

Subjects perform the classification for numerous images. A random subset of subjects then experience a system failure wherein they cannot classify images while the system resets. Subjects that experience the failure are randomly assigned to one of two treatments: 1) without operational transparency where they are told only how long to expect the failure to last or 2) with operational transparency they are told how long to expect the failure to last and the steps the system is performing to resolve the issue. Our results show that works who experience a failure and do not receive information about the steps the system in performing to get back to working order are less accurate at identifying flowers than those who experience the failure and receive operational transparency.

To investigate these questions, we conduct eight experiments (four at a large US university and four on Amazon Mechanical Turk) in which subjects are asked to perform traditional AI-training tasks. Table 1 shows an overview of the experiments. In the experiments, we first show that a system failure leads to an increase in errors in classification after the system recovers from failure and is fully operational. We also vary the payment scheme and the operational transparency to determine which managerial lever can reduce the negative impacts of the system failure. We explore several mechanisms that may help explain these results, such as attention, cognitive ability, engagement, effort, and confidence. Finally, we also explore the spillover of system failure of a focal task on unrelated future tasks. Together, our results demonstrate that the collaboration of human-AI systems deserves careful operational design and may have important implications for organizations in the midst of adopting AI systems.

Our results may contribute to the literature in several ways. First, we demonstrate that system failures can impact worker performance in AI-human interfaces. Second, we show that both operational transparency and performance-based pay can help to reduce the effects of system disruptions on the performance. Third, we provide evidence that the underlying mechanism through which the system failure impacts worker quality is the workers' confidence.

There may be several mechanisms why operational transparency affects workers dealing with failure in the workplace. For example, a worker may feel less satisfied or confident in his or her task. Thus, we also measure subjects' job satisfaction and confidence in the choices they made.

Our results so far show that satisfaction does not seem to explain the reduction in accuracy observed for subjects who experience a failure but do not receive operational transparency. However, in the first experiment, confidence levels are noticeably lower for students who experienced a system failure but did not receive operational transparency. Subjects who received operational transparency have confidence levels in line with those observed in students who did not experience a system failure. Finally, we test whether AI failure may impact the success rates of subjects in a different unrelated task to evaluate whether there is a "spillover" effect from the lack of transparency.

References

Buell, Ryan, Ethan Porter, Michael Norton. 2017. Surfacing the submerged state: Operational transparency increases trust in and engagement with government.

Buell, Ryan W, Michael I Norton. 2011. The labor illusion: How operational transparency increases perceived value. *Management Science* **57**(9) 1564–1579.

Ho, Chien-Ju, Aleksandrs Slivkins, Siddharth Suri, Jennifer Wortman Vaughan. 2015. Incentivizing high quality crowdwork. *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 419–429.

Olsen, Tava Lennon, Brian Tomlin. 2019. Industry 4.0: Opportunities and challenges for operations management. *Manufacturing & Service Operations Management* .

Wilson, H James, Paul R Daugherty. 2018. Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review* **96**(4) 114–123.