PACIS 2023 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

7-8-2023

Will Humans be Free-Riders? The Effects of Expectations for AI on Human-AI Team Performance

Mandie Liu
City University of Hong Kong, mandieliu2-c@my.cityu.edu.hk

Weiling Ke
Southern University of Science and Technology, kewl@sustech.edu.cn

David (Jingjun) Xu
City University of Hong Kong, davidxu@cityu.edu.hk

Will Humans be Free-Riders? The Effects of Expectations for AI on Human-AI Team Performance

Short Paper

Mandie Liu
City University of Hong Kong
Southern University of Science and Technology
mandieliu2-c@my.cityu.edu.hk

Weiling Ke
Southern University of Science and Technology
kewl@sustech.edu.cn

David (Jingjun) Xu
City University of Hong Kong
davidxu@cityu.edu.hk

Abstract

The failure of human-AI augmentation is a common problem that is usually believed to be highly related to poor AI design and human's inability to identify appropriate AI suggestions, but existing interventions like explainable AI were not effective to solve this problem. We propose that a crucial factor contributing to the failure of human-AI augmentation lies in the withholding of human effort. Moreover, high expectations for AI performance, which is generally positive for AI adoption, may undermine human-AI team performance by reducing human involvement in the task. Based on the Collective Effort Model (CEM), we explore how expectations for AI performance, perceive indispensability and task meaningfulness influence human effort and human-AI team performance. We plan to conduct laboratory experiments in image classification and idea generation to test our hypotheses. We expect to enhance the understanding of human-AI collaboration and the effects of social loafing effect in human-AI teams.

Keywords: Expectation for AI, human-AI team, human motivation

Introduction

AI is becoming increasingly professional in many human jobs, like diagnosis, customer service and translation (Schemmer et al., 2022). Nonetheless, despite AI's independent ability to perform tasks, many tasks still require human input, such as complementary knowledge, leadership and creativity (Fügener et al., 2022). Consequently, human-AI teams have gained widespread acceptance to achieve superior outcomes (Bansal et al., 2021). However, the failure of human-AI augmentation (i.e., the team performance exceeds the performance of both individual entities) has become a common problem, which typically results in poor human-AI team performance (Fügener et al., 2021; Liel & Zalmanson, 2020). Existing studies considered poor AI design and human's inability to identify appropriate AI suggestions are the main reasons for this failure (Hemmer et al., 2021). Thus, various improvements have been proposed to increase human-AI performance, like providing explainable AI (Zhang et al., 2020) and confidence score (Zhang et al., 2020), optimizing user interface (Buçinca et al., 2021), and improving human mental model (Bansal et al., 2019). However, empirical evidence suggested these interventions were not effective to improve human-AI team performance. In these studies, the underlying assumption is humans would be actively engaged in tasks and attempt to understand AI suggestions. However, we challenge this assumption and propose that human may withhold their effort in human-AI teamwork, thus previous improvements in AI design didn't play their due role. Indeed, existing research has indicated that humans were distanced from the work in human-AI collaboration tasks (Liel & Zalmanson, 2020; Buçinca et al., 2021; Vasconcelos et al., 2022), which will not only lead to low individual human performance but also blind acceptance and rejection of AI advice, thus resulting in less-than-optimal team performance. Therefore, we argue human motivation is crucial in enhancing human-AI team performance.

Human motivation is influenced by various factors, but little research has been placed on human perception of AI. Previous studies considered the high expectation for AI to be positive and have proposed various methods to increase it, as researchers believe high expectations will increase the likelihood of human-AI collaboration (Kocielnik et al., 2019; Zhang et al., 2021). However, we consider that high expectations for AI performance may undermine human motivation. In fact, human-AI collective work has formed AI and humans into a team. Previous studies in teamwork suggested that the expectations of co-worker performance significantly influence individual motivation in the collective task (Hart, et al., 2001). For example, Hüffmeier et al. (2013) suggested that high expectations for co-worker performance can lead to decreased individual effort, while low expectations can result in increased individual effort. However, our understanding of whether and how expectations of AI performance influence individual effort and performance in human-AI teams is still limited. Therefore, we propose the following research question:

RQ: How do the expectations for AI performance influence human-AI teamwork?

To answer this question, we applied the collective effort model (CEM) to build our theoretical framework and consider expectation comparison of human and AI performance, perceived indispensability and task meaningfulness as predictors of individual efforts in human-AI teamwork. We plan to conduct laboratory experiments to test our hypotheses in the context of image classification (disjunctive task) and brainstorming (additive task). Our study can enhance the understanding of individual motivation in human-AI teams and contribute to enhancing human-AI performance. Meanwhile, we can give suggestions for how to achieve high performance of human-AI teamwork.

Literature Review

Human-AI Teamwork

Although previous studies have focused on the collaboration of humans and AI, most studies explored how to increase the adoption of AI advice. Their results show that presenting AI advice usually increases human accuracy but will not outperform AI alone, which indicates the failure of augmentation of AI and humans. Researchers found various factors influencing the performance of human-AI teams. Poor AI design is regarded as a crucial factor. For example, Zhang et al. (2020) considered low transparency of AI suggestions undermines the collaboration of humans and AI. Bucinca et al. (2021) suggested low AI accuracy led to poor human-AI performance compared to humans and AI working alone. Researchers have found that human's inability to identify appropriate AI advice significantly influences human-AI performance. For example, Fügener et al. (2021) reported that AI advice will decrease human unique knowledge, which means humans are prone to blindly adopt AI advice. This overreliance leads to lower decision accuracy in human-AI teams (Bucinca et al., 2021; Liel & Zalmanson, 2020).

Many researchers have tried to improve the augmentation of humans and AI. The most common intervention is explainable AI. However, most studies have suggested that explainable AI does not necessarily enhance human-AI performance (Bansal et al., 2021; Liu et al., 2021). Explainable AI only increases the trust in AI advice, but it contributes little to human-AI performance (Alufaisan et al., 2021). Some researchers found that AI confidence score/certainty will help to trust calibration (Zhang et al., 2019; Fügener et al., 2021). But only enhancing trust calibration is not enough to enhance the whole performance of the human-AI team. All these interventions assumed that humans will actively engage in the task and carefully consider AI suggestions. Individual motivation has been supported as a crucial factor influencing team performance in management (Thomas & Velthouse, 1990) and psychology (Karau & Williams, 1993) research. Siemon and Wank (2021) proposed that the mere presence of AI-based teammates will lead to social loafing effect (i.e., humans will invest less effort in human-AI collaboration), but their pre-study results didn't support their hypothesis. In our study, we propose that the social loafing effect exists in human-AI actual collaboration. The reasons could be attributed to the expectations for AI performance, perceived indispensability, and task meaningfulness, beyond the mere presence of AI.

Expectations for AI Performance

Individual motivation is highly related to expectations for co-workers' performance (Williams & Karau, 1991; Hüffmeier et al., 2013). Previous studies have suggested that the expectation for co-worker performance will significantly influence individual effort. These expectations include co-workers' effort and ability. For example, people tend to pay for more effort in the tasks if their co-workers are expected to have low effort (Williams & Karau, 1991). Hart et al. (2001) also investigated how individuals pay for effort when the expectation for co-workers' ability and effort is different. Results showed that group members pay more effort when the partner had a low ability and pay less effort when the partner had a high ability. Researchers also found that individual effort is related to task demand (Hüffmeier et al., 2013) and task meaningfulness (Karau & Williams, 1993). Despite the significant impact of expectations for co-worker performance, the existing research only explored this effect in human-human collaboration. There is an increasing number of tasks completed by humans and AI. Given the characteristics of non-humanness and the different trust in AI, how individuals pay their effort in human-AI collaboration still needs to be explored.

Previous studies about the expectation for AI in human-AI collaboration focused on the positive effect of high expectations. For example, Zhang et al. (2021) consider high expectations for AI performance can lead to greater trust and confidence in AI, which can improve the probability of AI adoption. Meanwhile, some researchers applied expectation confirmation theory to explore how the expectation gap influences users' evaluations of AI systems (Riveiro & Thill, 2021). Our study proposes that high expectations may have negative effects on human-AI collaborations, and that high expectations would lead users to overly rely on their AI teammates thus leading to the loss of human complementary capabilities.

Collective Effort Model

The collective effort model (CEM), proposed by Karau & Williams (1993) introduced how group members pay their effort in collective tasks based on the expectancy × value framework (Wigfield & Eccles, 2000). In collective conditions, individuals work with other group members toward a single goal. It predicts the effort gains and loss in collective tasks based on three factors: expectancy, instrumentality and valence. Expectancy describes the expectation that high individual effort will lead to high individual performance. Instrumentality includes three expectations: the expectation that high individual performance will lead to high team performance, high team performance will lead to the desired outcomes and desired outcomes for the team will lead to desired individual outcomes. In our study, we assume that high team outcome predicts high team performance. As the partner is AI, the desired team outcome equals the desired individual outcome. Therefore, we only focused on the first expectation. Valence refers to the perceived value of the achievable outcome. According to the CEM, team members will exert their highest level of effort when all three factors are high (Karau & Williams, 1993). Several studies exploring social loafing have shown support for CEM (Karau & Williams, 1993). Social loafing is the reduction in motivation and effort when individuals work collectively compared to working individually (William & Karau, 1991). CEM provides a theoretical framework to understand under which conditions people engage in social loafing.

CEM has been widely used in human-human group work (Karau & Williams, 2014; Hüffmeier et al., 2013), and we apply this model to human-AI collaboration. While some people view AI as a tool or technology that won't evoke any social or emotional response (Seeber et al., 2020), team formation tends to make humans view AI as a teammate in human-AI collaboration (Nass et al., 1996; Siemon, 2022; Kim et al., 2022). Collaboration means the joint effort towards a common goal. AI usually provide complementary knowledge or gives solutions for humans in teamwork, which enables them to act like humans, not just automate certain tasks or processes. Rix (2022) provided evidence that the team formation of humans and AI will drive individuals to view AI as a teammate.

In our study, we aim to explore how expectations for AI performance influence individual motivation based on CEM. In CEM, expectancy only refers to the expectations for self-performance. However, other research found that expectations for co-workers also influence individual efforts because of social loafing (Karau & William, 1991). Therefore, expectations of self-performance are not sufficiently determined by individual efforts. In our study, we explore the effect of expectation comparison between self and AI performance. Instrumentality in CEM could be described by perceived indispensability, which refers to how individuals feel that their efforts were instrumental to team performance and it positively influences individual effort

(Hüffmeier et al., 2013). As for valence, extant literature usually utilized task meaningfulness to represent it, which refers to the value of the task goal or purpose (Hüffmeier et al., 2013).

Construct	Definition	Source
Expectation comparison between self and AI performance	The comparison of expectations for self and AI performance in case of high effort.	Hüffmeier et al. (2013)
Perceived indispensability for self-contribution	How individuals feel that their efforts were instrumental to team performance.	
Task meaningfulness	The value of the task goal or purpose.	

Table 1. Definition of Constructs

Hypotheses Development

We argue that expectation comparison for self and AI performance will influence individual motivation based on CEM. When individuals feel their AI teammate is superior to them in human-AI teamwork, individuals may withhold their effort. Previous studies suggested that working with high-ability partners tends to lead to a loss of individual effort based on the social loafing effect (Karau & Williams, 1991). Researchers found that humans usually view AI as human partners when working as a team (Rix, 2022). Individuals may consider it reasonable for AI to replace some human jobs, thus much effort is not required. Some researchers also found empirical evidence that humans will blindly follow AI advice when they feel AI has a higher ability (Fügener et al., 2021). Therefore, individuals are prone to rely on AI teammates and reduce their efforts when expecting a superior AI. When individuals feel their teammate is equally strong as them, they may pay more effort. People tend to be skeptical of AI decisions and are prone to work hard to prove themselves when they and AI both have strong expertise in this domain (Hemmer et al., 2021). They may perceive AI as a competitor who will damage their self-esteem. When individuals feel their AI teammate is inferior to them, they may spend more effort on human-AI teamwork compared to expecting an equally strong AI. People tend to have lower trust in such type of AI. Meanwhile, previous studies suggested that people who expect to work with an inferior co-worker will lead the increased effort in teamwork (Karau & William, 1991). In human-AI teams, individuals may pay more effort out of responsibility and to avoid failure.

H1: Expectation comparison for self and AI performance will influence human effort in human-AI tasks, so that (a) compared to working individually, people expecting to work with a superior AI will invest less effort, (b) people expecting to work with an equally strong (vs. superior) AI will invest more effort, (c) people expecting to work with an inferior (vs. equally strong) AI will invest more effort.

Based on CEM, perceived indispensability will positively influence individual effort in group work. This is because people are often motivated by personal responsibility and the desire to maintain a positive self-image (Hüffmeier et al., 2013). In a human-AI team, if individuals perceive their contribution as essential and irreplaceable to the success of the team, they may feel a greater sense of obligation to pay extra effort and perform at a high level.

H2: Perceived indispensability for self-contributions will positively influence human effort in human-AI teams.

CEM also proposes valence is an important factor in determining individual effort. High task meaningfulness can increase intrinsic motivation (Hunton et al., 1997) and collaboration of team members (Karau & William, 1991). In a human-AI team, if the task meaningfulness is high, people are more likely to feel satisfied and fulfilled and to strive toward team success. Meanwhile, humans may have more willingness to consider AI advice, which requires them to pay more effort in the collaboration with AI.

H3: Task meaningfulness will positively influence human effort in human-AI teams.

Previous studies have suggested that high effort will lead to high performance when the ability is high (Gatewood et al., 2022). When individuals have the necessary skills and expertise, paying more effort can maximize their performance. In human-AI teams, both humans and AI usually have their unique knowledge. Therefore, enough human effort is the basis of the team's success, as AI will already provide maximized effort to complete their task. Additionally, group dynamics such as communication and trust will also influence team performance (Driskell & Salas, 2006). When humans are actively involved in the task, they can better understand the decision made by AI and integrate AI suggestions with their own ideas, which can effectively prevent blind adherence or rejection of AI decisions.

H4: Human effort in human-AI teams will positively influence the performance of human-AI teams.

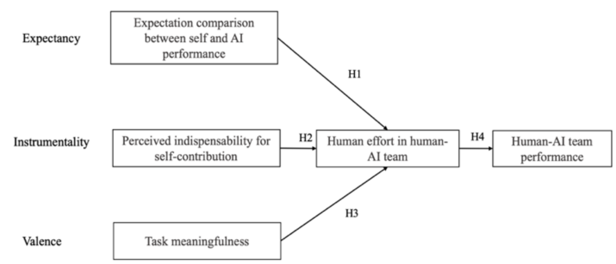

Figure 1. Research Model

Research Design

We plan to conduct laboratory experiments with a 4 (work individually, expectation comparison: AI is superior/inferior/equally strong to individual) × 2 (perceived indispensability: high/low) × 2 (task meaningfulness: high/low) between-subjects design to test our hypotheses. We will manipulate expectation and task meaningfulness in the description of the experiment instruction and manipulate perceived indispensability by different tasks.

Treatment

Expectation comparison

For the treatment of expectation comparison for self and AI performance, we will provide different descriptions of the AI before they see the task introduction. As our tasks don't need any professional knowledge and skills, expectations for self-performance are not expected to vary much. Therefore, we provide different descriptions for AI ability, which will determine the expectation comparison of self- and AI performance. Although the description of AI is different, the accuracy of AI will be the same. We will first confirm that participants have read our description and then ask them to answer questions about their expectations for self-performance and AI performance before they see the task content.

Task meaningfulness

We will manipulate the task meaningfulness in task instruction. According to the definition of task meaningfulness, high task meaningfulness usually requires people's work to be recognized by others and makes them understand their work can be related to the greater good (Chandler & Kapelner, 2013). In our study, participants in the group of high task meaningfulness will be informed of the purpose of the experiment and the importance of the task, and they will be thanked for participating (Chandler & Kapelner, 2013; Wang et al., 2022). The instruction will clearly inform them that this experiment aims to examine the quality of intelligence in adults (Karau & William, 1991), and the experimental data will contribute to the development of artificial intelligence (Wang et al., 2022). Apart from that, we will also enhance external motivation by telling participants that they will receive 2 dollars if they perform well. In low meaningful conditions, we won't give any reason for their task.

Perceived indispensability

We will manipulate perceived indispensability by task demand. Previous studies indicated that different task demands influence group members' perceived indispensability (Karau & William, 1991; Hüffmeier et al., 2013). The task can be divided into the disjunctive task, additive task and conjunctive task based on task demand. In the disjunctive task, all team members have the ability to complete the task independently, but the whole performance is determined by the best individual performance. Image classification and choosing a new location for office are typical examples of disjunctive tasks. In additive tasks, although all individuals have the ability to complete the task, it's hard to get a superior outcome. Thus, the sum of humans and AI's work contribute to the final team performance. Brainstorming is a typical example of an additive task. In conjunctive tasks, individuals in teams need to do tasks one by one to complete the whole task, and the whole performance is determined by the weakest individual performance. Assembly line work by humans and AI belongs to conjunctive tasks.

But our study aims to explore the effects of expectation comparison for self- and AI performance, which requires AI and humans to do the same task. Therefore, we only conduct experiments in disjunctive and additive tasks. Previous studies suggested that indispensability perception can be induced by how individual contributions relate to team success (Hüffmeier et al., 2013). In disjunctive tasks, both humans and AI can lead to team success alone, which means human work can be replaced by AI. Thus, in this case, individuals' perceived indispensability is low. In additive tasks, humans can make some unique contributions, as the sum of the members' contributions determines the team's success. Thus, individuals' perceived indispensability is high in additive tasks.

Disjunctive task: image classification (low indispensability) We choose image classification as an experimental context for the following reasons: first, image classification is a task that AI and humans can perform well alone, and the whole performance is determined by the final choice, which makes it a good example of a disjunctive task. Also, well-designed AI can replace human work in image classification. Thus, the subjects' perceived indispensability will be low in this case. Second, the AI performance is determined by its design. Thus, we can manipulate it through AI descriptions. Third, image is a generic task that all humans can complete without any professional skills or knowledge.

In the task of image classification, the participants will be required to assign a focal image (e.g., the image of a flower) to one of 10 possible image classes. Each of these classes will be accompanied by a class name (e.g., "Sunflower" or "Calendula") and ten sample images representing that class, similar to Fügener et al. (2021). Each participant will engage in the same ten sets of the image classification task. These 10 tasks are not straightforward (e.g., the color of flowers in all classes is the same). Participants need to observe these images carefully, which means they must pay some effort into their tasks. And they will be informed that the task performance is determined by their classification accuracy.

Additive task: idea generation (high indispensability) We choose idea generation as the experimental context for the additive task. We will ask subjects to come up with as many uses as possible for a knife for the following reasons: first, the team performance is determined by the sum of ideas generated by AI and subjects, which is a good example of an additive task. Although AI can perform well in idea generation, humans can come up with many innovative ideas, which means subjects can make some unique contributions. Thus, the perceived indispensability in idea generation is higher than in image classification. Second, idea generation is a generic task that doesn't need any professional knowledge or skills. Third, AI performance can be determined by the design (e.g., database, searchability).

Participants will be given 10 minutes to brainstorm and write down their ideas for the use of a knife. They can submit their answers whenever they feel they finished the task. In the group "work individually", participants will be told that they need to come up as many uses as possible on their own. In the group with AI, participants will be told that an AI partner will be working on this task and it doesn't matter if they come up with the same ideas as the AI. They will be informed that the task performance is evaluated based on the combined performance of both themselves and the AI partner.

Treatment	Manipulation	Manipulation check
Expectations comparison	Superior: "The advice is from one of the best-performing AIs designed by Google. It performs well in previous tasks." Equally strong: "The advice is from an AI designed by a small company. It performs normally in previous tasks." Inferior: "The advice is from an AI designed by an undergraduate student for a class assignment. It performs not satisfactorily in previous tasks."	We will 1) ask questions about the designer of the AI. 2) compare the expectation for self- and AI performance.
Task meaningfulness	Low: "You will be working on 10 image classification tasks."	We will 1) ask questions about the

	High: "Thanks for participating in this task! You will be working on 10 image classification tasks. The purpose of this experiment is to examine a quality that is thought to be highly correlated with intelligence in adults. Our experimental data will also contribute to the development of artificial intelligence. You can get 2 dollars if you perform well. "	purpose of the study in "high task meaningfulness" group 2) measure task meaningfulness.
Perceived indispensability	Low: image classification High: idea generation	Measuring perceived indispensability

Table 2. Manipulation for Treatment

Measurement

Expectations for self-performance and AI performance will be measured before starting the task. We will separately measure the two expectations and compare them. Perceived indispensability and task meaningfulness will be measured at the end of the experiment to avoid the demand effect. Human effort will be measured based on participants' behavior. We will adopt different measurements for human efforts in image classification and idea generation. As image classification is similar to a student test that has some items to complete one by one, we plan to employ the student engagement index in the test-taking process to measure effort (Wise, 2015). For the idea generation, we will use a response time scale to measure effort due to the time limitation imposed on this task (Shroyer et al., 2018).

For human effort measurement in image classification, response time and users' attention are two important indicators of the student engagement index. We will adopt response time effort (RTE) as the time-based metric for human effort in image classification. Initially, we will record the start and finish time for each classification task and calculate the average response time of each task. Subsequently, a time threshold will be established for each task. To identify non-effortful participants, we will employ the normative threshold time (NT20) proposed by Wise (2015), which categorizes participants with a response time less than 20% of the average response time as non-effortful. For instance, if the average response time for classification task 1 is 20 seconds, participants with a response time of less than 4 seconds will be considered non-effortful. Finally, we will calculate the proportion of classification tasks completed with effort. For example, since there are a total of 10 classification tasks, if a participant completes 2 tasks effortlessly, the RTE score will be 0.8. Apart from response time, attention will be measured by evaluating participants' ability to recall task content and correctly answer five questions regarding important details (e.g., petal shape). The attention index will be derived from the proportion of correct answers to these questions. In idea generation, human effort measurement will begin by calculating the difference between the shortest and longest response time. This time difference will then be divided into 10 equal intervals and assigned codes ranging from 0.1 to 1.0 (Shroyer et al., 2018).

Task performance will be measured by accuracy for image classification and the quantity and quality of ideas for idea generation. The quantity of ideas will be determined by counting the total number of ideas generated by participants. The quality of each idea will be evaluated based on criteria of creativity and feasibility (Potter & Balthazard, 2004). ANCOVA analysis will be adopted to test hypotheses. We will also examine the potential interaction effects of the three factors.

Construct	Sampled measurement items	Source
Expectation comparison	If you/AI pay high effort in the task, how well do you expect you/AI to perform on this task?	Wigfield & Eccles (2000)
Perceived indispensability	I think I can make unique contributions to team success.	Kerr & Hertel (2011)
Task meaningfulness	I feel that the work I do on the job is valuable.	May et al. (2004)
Human effort	For image classification, we will measure response time effort and attention; for idea generation, we will measure response time scale.	Shernoff et al. (2003); Wise (2015)

Table 3. Measurement items

Expected Contributions

Our study is expected to make several theoretical contributions. First, we aim to enhance the understanding of human-AI performance from the perspective of human motivation. Previous studies have focused on how to increase AI adoption and found that increasing AI adoption has little effect on enhancing human-AI performance. We propose that human motivation in human-AI collective tasks is a crucial predictor of performance by increasing appropriate reliance on AI. Second, we further explore the social loafing effect in human-AI collaboration by CEM. Siemon and Wank (2021) proposed that the mere presence of AI-based teammates leads to social loafing in human-AI teams but the results didn't support their hypothesis. We consider expectations for AI performance, task demand, and task meaningfulness all influence the social loafing effect in human-AI collaboration settings, not just the mere presence of AI. Third, we extend CEM into the collaboration of humans and AI. Our study can provide evidence for CEM on human-computer interactions, which has been widely studied in human-human interaction. Practically, we expect to give suggestions for managers to better manage human-AI teams.

Acknowledgements

This work was partially supported by Research Grants at the City University of Hong Kong (Grant Nos. 7005595 and 9680306).

References

Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., ... & Weld, D. (2021). Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-16).

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1-21.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, *90*, 123-133.

Driskell, J. E., & Salas, E. (2006). Groupware, Group Dynamics, and Team Performance.

Fügener, A., Grahl, J., Gupta, A., & Ketter, W. (2021). Will Humans-in-the-Loop Become Borgs? Merits and Pitfalls of Working with AI. Management Information Systems Quarterly (MISQ)-Vol, 45.

Gatewood, E. J., Shaver, K. G., Powers, J. B., & Gartner, W. B. (2002). Entrepreneurial expectancy, task effort, and performance. *Entrepreneurship theory and practice*, *27*(2), 187-206.

Hart, J. W., Bridgett, D. J., & Karau, S. J. (2001). Coworker ability and effort as determinants of individual effort on a collective task. *Group Dynamics: Theory, Research, and Practice*, *5*(3), 181.

Hemmer, P., Schemmer, M., Vössing, M., & Kühl, N. (2021). Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. *PACIS*, 78.

Hunton, J. E., & Price, K. H. (1997). Effects of the user participation process and task meaningfulness on key information system outcomes. *Management Science*, *43*(6), 797-812.

Hüffmeier, J., Dietrich, H., & Hertel, G. (2013). Effort intentions in teams: Effects of task type and teammate performance. *Small Group Research*, *44*(1), 62-88.

Juravle, G., Boudouraki, A., Terziyska, M., & Rezlescu, C. (2020). Trust in artificial intelligence for medical diagnoses. *Progress in brain research*, *253*, 263-282.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. Journal of Personality and Social Psychology, 65, 681-706. doi:10.1037/0022-3514.65.4.681

Karau, S. J., Markus, M. J., & Williams, K. D. (2000). On the elusive search for motivation gains in groups: Insights from the collective effort model. *Zeitschrift für Sozialpsychologie*, *31*, 179-190. doi:10.1024//0044-3514.31.4.179

Karau, S. J., & Williams, K. D. (2014). Understanding individual motivation in groups: The collective effort model. In *Groups at work* (pp. 127-156). Psychology Press.

Kerr, N. L., & Hertel, G. (2011). The Köhler group motivation gain: How to motivate the "weak links" in a group. *Social and Personality Psychology Compass, 5, 43-55.doi:10.1111/j.1751-9004.2010.00333.x*

Kim, J., Merrill Jr, K., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383.

Kocielnik, R., Amershi, S., & Bennett, P. N. (2019, May). Will you accept an imperfect ai? exploring designs for adjusting end-user expectations of ai systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Liel, Y., & Zalmanson, L. (2020). What If an AI Told You That 2+ 2 Is 5? Conformity to Algorithmic Recommendations. In *ICIS*.

Liu, H., Lai, V., & Tan, C. (2021). Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-45.

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of occupational and organizational psychology*, *77*(1), 11-37.

Potter, R. E., & Balthazard, P. (2004). The role of individual memory and attention processes during electronic brainstorming. *Mis Quarterly*, 621-643.

Price, K. H., Harrison, D. A., & Gavin, J. H. (2006). Withholding inputs in team contexts: Member composition, interaction processes, evaluation structure, and social loafing. *Journal of Applied Psychology*, *91*(6), 1375.

Riveiro, M., & Thill, S. (2021). “That's (not) the output I expected!” On the role of end user expectations in creating explanations of AI systems. *Artificial Intelligence*, *298*, 103507.

Rix, J. (2022). From Tools to Teammates: Conceptualizing Humans’ Perception of Machines as Teammates with a Systematic Literature Review.

Thomas, K. W., & Velthouse, B. A. (1990). Cognitive elements of empowerment: An “interpretive” model of intrinsic task motivation. Academy of Management Review, 15(4), 666–681.

Schemmer, M., Hemmer, P., Kühl, N., Benz, C., & Satzger, G. (2022). Should I Follow AI-based Advice? Measuring Appropriate Reliance in Human-AI Decision-Making. arXiv preprint arXiv:2204.06916.

Seeber, I., Bittner, E., Briggs, R. O., De Vreede, T., De Vreede, G. J., Elkins, A., ... & Söllner, M. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, *57*(2), 103174.

Seta, J. J., Seta, C. E., & Donaldson, S. (1991). The impact of comparison processes on coactors' frustration and willingness to expend effort. *Personality and Social Psychology Bulletin*, *17*(5), 560-568.

Siemon, D. (2022). Elaborating team roles for artificial intelligence-based teammates in human-AI collaboration. *Group Decision and Negotiation*, *31*(5), 871-912.

Siemon, D., & Wank, F. (2021). Collaboration With AI-Based Teammates-Evaluation of the Social Loafing Effect. In *PACIS* (p. 146).

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School psychology quarterly*, *18*(2), 158.

Shroyer, K., Lovins, T., Turns, J., Cardella, M. E., & Atman, C. J. (2018). Timescales and ideaspace: An examination of idea generation in design practice. *Design Studies*, *57*, 9-36.

Steiner, I. D. (1972). *Group process and productivity* (pp. 96-98). New York: Academic press.

Wang, L., Zhang, X., & Meng, L. (2022). Work meaningfulness disclosure and persistent behavior: Level of action identification as a moderator. *PsyCh Journal*, *11*(1), 77-84.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, *25*(1), 68-81.

Williams, K. D., & Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of personality and social psychology*, *61*(4), 570.

Wise, S. L. (2015). Effort analysis: Individual score validation of achievement test data. *Applied Measurement in Education*, *28*(3), 237-252.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M., & Krishna, R. (2022). Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *arXiv preprint arXiv:2212.06823*.

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 295-305).

Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1-25.