Rebecka C. Ångström

DANCING WITH THE DYNAMIC MACHINE

ESSAYS ON IMPLEMENTATION PRACTICES, TRUST DYNAMICS, AND IDEA EVALUATION IN AND AROUND THE USE OF ARTIFICIAL INTELLIGENCE IN ORGANIZATIONS

STOCKHOLM SCHOOL OF ECONOMICS

DANCING WITH THE DYNAMIC MACHINE

Artificial Intelligence (AI) is advancing its position in organizations by performing tasks historically perceived as exclusive to humans. As AI becomes more commonplace in the work environment, there is an increasing need to understand the implications of integrating AI into the fabric of organizations.

This thesis investigates how AI-related dynamics are manifested in the organization and their impact on employee trust and distrust in AI. The thesis consists of three articles, each based on a unique data set, including a multinational survey, a longitudinal case study, and a field experiment. Together, the articles show that AI will not only induce continuous transformation of the organization but can also generate persistent uncertainty amongst employees. Such uncertainty can then be amplified by employees' and managers' limited understanding of AI functionality, resulting in distrust. The results also show that human biases, a challenge expected to be addressed using AI, manifest differently than commonly believed. Such misconceptions can create unrealistic expectations on AI, further contributing to employee uncertainty.

The thesis ends with a call to increase our general knowledge regarding AI and its reliance on organizational data, followed by three suggestions for future research on the implications of integrating AI in organizations.

REBECKA C. ÅNGSTRÖM is a researcher at the Department of Innovation, Entrepreneurship, and Technology, House of Innovation, at the Stockholm School of Economics and Principal Researcher at Ericsson Research.

ISBN 978-91-7731-277-2

DOCTORAL DISSERTATION IN BUSINESS ADMINISTRATION
STOCKHOLM SCHOOL OF ECONOMICS, SWEDEN 2023

Dancing with the Dynamic Machine

Essays on Implementation Practices,
Trust Dynamics, and Idea Evaluation in and around
the use of Artificial Intelligence in Organizations

Rebecka Cedering Ångström

Akademisk avhandling

som för avläggande av ekonomie doktorsexamen
vid Handelshögskolan i Stockholm
framläggs för offentlig granskning
fredagen den 29 september 2023 kl. 10.15,
i Lecture Hall, Handelshögskolan,
Norrtullsgatan 2, Stockholm

Dancing with the Dynamic Machine

Essays on Implementation Practices, Trust Dynamics, and Idea Evaluation in and Around the Use of Artificial Intelligence in Organizations

Dancing with the Dynamic Machine

Essays on Implementation Practices, Trust Dynamics, and Idea Evaluation in and Around the Use of Artificial Intelligence in Organizations

Rebecka C. Ångström

Dissertation for the Degree of Doctor of Philosophy, Ph.D.,
in Business Administration
Stockholm School of Economics, 2023

Dancing with the Dynamic Machine: Essays on Implementation Practices, Trust Dynamics, and Idea Evaluation in and Around the Use of Artificial Intelligence in Organizations

ISBN 978-91-7731-277-2 (printed)
ISBN 978-91-7731-278-9 (pdf)

Front cover illustration:
© Rebecka Cedering Ångström, 2023. Created with the assistance of DALL·E 2

Back cover photo:
Rebecka Cedering Ångström, 2023

Printed by:
BrandFactory, Stockholm, 2023

Keywords: Artificial Intelligence, AI-driven organization, Technology implementation, Implementation challenges, Trust, Distrust, Trust transference, Idea evaluation, Bias

To
Rakel and Edith,

For all the doors you will open

Foreword

This volume is the result of a research project carried out at the Department of Entrepreneurship, Innovation and Technology at the Stockholm School of Economics (SSE).

The volume is submitted as a doctoral thesis at SSE. In keeping with the policies of SSE, the author has been entirely free to conduct and present her research in the manner of her choosing as an expression of her own ideas.

SSE is grateful for the financial support provided by Ericsson Research, which has made it possible to carry out the project.

Göran Lindqvist

Director of Research
Stockholm School of Economics

Magnus Mähring

Professor and Head of the
Department of Entrepreneurship,
Innovation and Technology

Acknowledgements

Writing a thesis is an ever-changing experience, including moments of ease, where I solved tasks almost effortlessly, and challenging times, which demanded unwavering dedication and grit to succeed. Honestly, the latter experiences outnumber moments of ease by far. However, throughout this process, I have received the support of colleagues, friends, and family, which all played vital roles in my learning, progress, and perseverance. Hence, I want to express my heartfelt appreciation to all of you who have been by my side during this incredible voyage.

First, I want to thank Magnus Mähring and Martin W. Wallin; you are both excellent researchers guiding me through this process, sharing your expertise and much laughter along the way. Magnus, I am truly grateful for your ability to consistently provide clarity and calmness during the many moments I found bewildering and stressful. You have taught me a lot. Martin, thank you for always being available and open to discussing big and small reflections. I have come to appreciate you like a friend. This is the way! I also thank my other co-authors mentioned in no particular order, Linus Dahlander, Eivor Oborn, Michael Barrett, Arne Thomas, and Michael Björn, for allowing me to work with you. I have learned so much from our work together. I further want to express my gratitude to the academic community and all the researchers at the House of Innovation (HOI); thank you for the discussions, ideas, and feedback over the years. I especially thank Anna Essén, Elmira van den Broek, Roberto Verganti, and Sebastian Krakowski for our inspiring discussions regarding AI.

I also like to thank the IMIT research school. Beyond Martin W. Wallin, I thank Joakim Björkdahl, Karl Wennberg, Martin Sköld, and Tobias Fredberg for sharing your expertise and support throughout my studies. From the academic community, I also want to thank Frida Pemer and Lisen

Selander for taking on the role of my opponent during different defenses. Thank you for moving my research and thinking forward. Then there are the ones that make the world go round; Thank you for your support and efforts, Kerstin Wedin, Marie Wahlström, Lee Malm, Agneta Carlin, all from Stockholm School of Economics, and Carina Blomkvist and Maria Christiansen from IMIT.

Finally, to my fellow PhD-students at HOI, I am genuinely grateful for having made this journey with all of you. The memories we share and the stories we can tell. You are the best.

I would not have been able to pursue my Ph.D. studies without Ericsson Research and Magnus Frodigh, Pernilla Jonsson, Anders Erlandson, and Anders Caspar. Thank you for providing this opportunity and making this journey possible. I look forward to sharing what I have learned with you and our community. I especially thank you, Pernilla and Anders, for your unwavering support and belief in me. I also want to express my sincere gratitude to my research partner in crime since many years back, Michael Björn; How I love to agree and disagree with you. Your support during my Ph.D. studies has been vital to me. Finally, I want to thank my wonderful colleagues at Ericsson for your encouragement and support.

I also take this opportunity to send a special thanks to my network of AI professionals and experts, including my fellow founders of AddAI – Thank you for keeping my curiosity and research interest in AI alive.

One of my most significant sources of positive energy has been my friends. Alina, my rock and best friend, your support, wits, and madness are always invaluable to me. To my wonderful friends in the zombie-skumpa-klubb, for all the wine, laughter, and suggestions on books I never had the time to read. You are always just a text message away. To my friends identifying as Mickelkåren, Smurfer, and general geeks – Thank you for all the wonderful adventures, carrying of barrels in the night, dinners, and laughter we shared over the past five years. Each experience provided me with the energy to keep moving forward. Finally, I am deeply grateful to my family. My mom for believing in me and always being ready to help when schedules collide. To my dad for continuously inspiring me to pursue more knowledge. To my daughters, Rakel and Edith, for being the most extraordinary people I know. I am astonished and inspired by your creativity and originality every day. And especially, Johannes, thank you for being by my side throughout this journey, keeping me grounded, fed, and sane – I love you with all my heart.

In the spirit of working with AI, the cover image of this dissertation is the product of a collaboration between Dall-E 2 and myself, where Dall-E 2 generated the illustration on my prompt "Many humans dancing with many dynamic machines, graphical design." The fact is that I enjoyed the experience of collaborating with Dall-E 2 so much that I accidentally spent the better part of a day generating almost 500 images before running out of credits and having to pick a favorite. I chose this illustration as it, at first glance, looks like human marionettes dancing in front of a giant cogwheel, but when you look closer, you see that the human-like shapes are not at all constrained by the strings.

Stockholm, August 3, 2023

Rebecka C. Ångström

Contents

Chapter 1

Introducing AI

In 2015 I accidentally killed the smartphone. Working at Ericsson, my colleague Michael and I had just published our new consumer technology trend report. One of the trends highlighted that early adopters worldwide were interested in communicating with their home appliances by talking to an AI instead of fiddling around with a smartphone app. Over half of the respondents also answered that they believed the smartphone would be a thing of the past within five years. Releasing the report, TT, a Swedish news agency, interviewed me about the trends. They thought the AI vs. smartphone trend was so interesting that they titled the news article "The smartphone will be dead in five years." Following the article's publication, the news that Ericsson had predicted the smartphone's death spread like wildfire worldwide.[1] To our disappointment, it was not the (at the time) novel usage of AI that had caught readers' interest, but the idea that our smartphone era might come to an end (and that it was Ericsson, a mobile network provider, that predicted it). The learning I bring from this incident is that it is difficult to discuss what lies beyond a prevailing paradigm, in this case, the smartphone.

Today, following the big breakthrough of generative AI, such as ChatGPT, I do not believe the AI vs. smartphone trend would have made the same headlines or had the same reach. Our news and social feeds are

---

[1] For examples of still available online news articles that republished TT's interview, see
www.cnbc.com/2015/12/09/people-think-the-smartphone-will-be-dead-in-5-years-ericsson.html
www.svd.se/a/4a6a8d05-daf5-445f-a747-78cc739b9395/smarta-telefoner-spas-do-ut-inom-fem-ar
https://guardian.ng/technology/smartphones-to-become-extinct-by-2020/

already bursting with ideas, tips, and tricks on utilizing new AI applications. Although I am thrilled that one of my favorite topics is now legit at the dinner table, I am concerned that our understanding of AI is still relatively shallow. Public discourse quickly rallies around the latest AI sensation or a distant sci-fi future, often fueled by fear. Although I do not expect to make headlines with this dissertation, I hope to contribute to a healthier discussion of AI and how it influences our organizations and work.

Artificial Intelligence (AI) is advancing its position in organizations by performing tasks historically perceived as exclusive to humans. From overtaking human manual labor, machines are now able to perform human cognitive and non-routine work (e.g., Brynjolfsson & Mitchell, 2017; Frey & Osborne, 2017) such as writing news articles (Carlson, 2015; GPT-3, 2020; Rai et al., 2019), predicting where the next crime will occur (Brayne, 2017; Shapiro, 2017; Waardenburg et al., 2022), driving cars, and directing city traffic (Baker, 2018; Davies, 2018). Naturally, this has resulted in practitioners and scholars debating how AI technology will impact human work and organizations. A central discourse has focused on whether machines will come to augment humans in their labor or automate processes that render human occupations redundant (Fleming, 2019; Frey & Osborne, 2017; Raisch & Krakowski, 2020). Others have explored how the introduction of AI will alter work; for instance, decision-making, expertise, and identity (Berente et al., 2021; Faraj et al., 2018; Vaast & Pinsonneault, 2021; von Krogh, 2018). Scholars have also identified trust in AI as critical for successfully integrating AI into an organizational context (Glikson & Woolley, 2020; Leonardi et al., 2022; Lockey et al., 2021) and found that a key factor for trust is the technology's impact on work (Gillespie et al., 2021). Researchers have also demonstrated that when employees do not trust AI to be to their benefit, they may develop strategies to resist the integration of AI in their workplace (Brayne & Christin, 2021; Christin, 2017; Kellogg, 2020).

Although many interesting and important questions have been raised and explored in prior research, some aspects of how AI influences the ongoing transformation of organizations are often overlooked. AI embodies three dynamics that drive transformational change in organizations. These dynamics are an effect of AI's inherent data-driven adaptability and the necessity of dynamic responses from organizations. First, as AI draws from data to generate predictions, organizations can adapt their processes from a reactive to a proactive orientation (Agrawal et al., 2018; Iansiti & Lakhani, 2020). Second, as AI is a general-purpose technology (Brynjolfsson & McAfee, 2017), becoming an AI-driven organization goes beyond developing and implementing a single AI application.[2] Instead, an organization will integrate numerous AI applications across the organization for various purposes (Agrawal et al., 2018; Iansiti & Lakhani, 2020). In this context, AI applications can generate new insights, thus creating opportunities for further AI applications. Third, AI has an inherent dynamic. In contrast to traditional IT software, AI can continuously learn and improve its accuracy over time as it is used or exposed to more data (Agrawal et al., 2018; Faraj et al., 2018). Combining these three dynamics reveals that AI technology can drive transformational change. However, transformational change within organizations can generate uncertainty and challenge employee trust (Gustafsson et al., 2021; Sørensen et al., 2011). Furthermore, AI's impact on work is a key factor for trust in AI (Gillespie et al., 2021). Thus, to better understand how we as humans come to trust AI in our work, we must consider how AI's dynamics are manifested and unfold as AI becomes integrated into organizations.

My thesis explores the dynamics of AI and the uncertainties it generates amongst employees. The goal is not to create an exhaustive description of phenomena but to explore how AI-related dynamics are manifested in the organization and the implications for employee trust in AI. I pursue this question through the three studies in the thesis, each with a unique dataset: a multi-national survey, a longitudinal case study, and a field experiment. In Paper 1, a multi-national survey, my co-authors and I targeted decision-makers in organizations currently implementing AI in their organizations. We aimed to better understand the challenges encountered during AI implementation and the strategies to overcome them. We reveal that, when implementing AI, challenges persist over time, even as organizations become more experienced. Paper 1 also shows that employee fear and resistance towards AI are some of the most common challenges across organizations. In Paper 2, a longitudinal case study, my co-authors and I followed a data science team

[2] I make a distinction between *implementing* AI, refering to the activities an organization undertakes to introduce a new technology (Myers, 1995), and the *integration* of AI into an organizational context where AI becomes a part of an AI and data-driven organization (Iansiti & Lakhani, 2020).

developing AI applications to help the corporate initiative become AI-driven. We show that employee distrust towards AI is not limited to a single AI application. We also reveal that AI's inherent dynamic, continuously learning from data, can be a source of distrust if the user lacks an understanding of the phenomenon. In Paper 3, a field experiment, my co-authors and I sought evidence of bias in idea evaluation. Bias is a known challenge for trustworthy AI (Maslej et al., 2023), and yet, as we show in our paper, it remains an area that we still do not fully understand, as our field experiment revealed a null result where we expected to find bias. The lack of bias is a vital insight into further understanding how human bias may impact AI accuracy. Combining the three papers and datasets, I offer three conclusions. First, becoming AI-driven is a continuous transformation where challenges persist over time, and employee distrust towards AI is one of the most common challenges organizations encounter. Second, the continuous development of numerous AI applications results in an ongoing interplay between social and technical trust referents, which can result in vicious distrust cycles. Third, the lack of understanding of AI's inherent dynamic, continuously learning from data, can lead to unrealistic expectations of AI and distrust towards the technology.

The rest of the thesis is structured as follows: In Chapter 2, I provide an account of relevant literature related to the focal phenomenon and theory background. In Chapter 3, I describe the research design, empirical context, and methods used in the three papers. I have also included a section on my position as an industrial Ph.D. student. Following the research design, in Chapter 4, I summarize each paper's findings and contributions relevant to the thesis research questions in a findings section. For Papers 1 and 3, I have also included additional data from their respective datasets that I found relevant to this thesis's overarching aim. In Chapter 5, I discuss the combined results of the three papers and their datasets in relation to the relevant literature. I also present ideas on potential future research built on AI as a dynamic machine. Lastly, in Chapter 6, I present the three papers.

A caveat regarding the order of the three papers: as the studies were conducted in parallel, I have decided not to present them chronologically. Instead, they follow a thematic approach, starting with a broader scope on implementing AI in organizations (Paper 1), followed by a closer look at how distrust emerges during integrating AI into the organizational context (Paper

2) and ending with a closer look at bias (Paper 3), which is a source of unreliable AI and a challenge to trustworthiness.

Chapter 2

Theoretical framework

In this chapter, I describe the dynamic aspects of AI and share the most relevant discourse for this thesis around AI in the organization and trust in AI. I conclude this section with an overview of the literature on human and AI bias and its relevance to AI trustworthiness. However, I begin with a short description and definition of AI.

What is AI?

Driven by computational power, progress in data science, and the availability of large datasets, AI comprises a group of technologies, such as machine learning, pattern recognition, computer vision, and natural language processing. The definitions of AI are numerous, and where some focus only on its technical capabilities (OECD), others include references to human intelligence (Rai et al., 2019), expectations of future technology (Berente et al., 2021), and even industry formation, such as politics, labor, and culture (Crawford, 2021). One of the reasons for this multitude of definitions is that the term 'AI' is somewhat of a moving target, whereby as soon as an AI technology becomes mundane, it is no longer perceived as AI (Berente et al., 2021; McCorduck, 2004). In this thesis, I follow Faraj et al. (2018) in using a single term, AI, to refer to "an emergent family of technologies that build on machine learning, computation, and statistical techniques, as well as rely on large datasets to generate responses, classifications, or dynamic predictions that resemble those of a knowledge worker" (Faraj et al., 2018, p. 62). This definition not only emphasizes that AI is a group of technologies but is generous about the type of technologies that it includes, such as machine learning and statistical techniques. Such an inclusive definition is helpful when dealing with a moving target. If anything, the definition could have also included more advanced subgroups of AI technologies, such as large language models. Another characteristic of this definition is that it highlights AI dependency on data. It is its ability to digest and analyze vast amounts of data in order to identify patterns, predict outcomes, and propose proactive solutions that differentiate AI from traditional IT software (Agrawal et al., 2018; Faraj et al., 2022). Moreover, it is the data enabling the three dynamics of AI that is relevant for this thesis (see below for a further description of the dynamic aspects). Lastly, the definition also includes references to what AI can generate and the range of AI applicability.

In this thesis, I use the term 'AI' when describing the technology in general and 'AI application' when mentioning a specific AI model. Occasionally, especially in Paper 2, I use the term 'algorithm' interchangeably with 'AI application.' In this thesis, I also use the term 'AI-driven' to refer to organizations undergoing major transformations striving to integrate AI across the organization to support and automate various processes (Agrawal et al., 2018; Iansiti & Lakhani, 2020).

The dynamics of AI

Enabled by data, AI embodies three dynamics that grant it transformational power. These dynamics are an effect of AI's inherent adaptability, learning from new data, and the necessity for dynamic responses as organizations adapt to AI. First, drawing from data, AI predictions enable organizations to shift from a reactive to a proactive operation (Agrawal et al., 2018). This shift means that instead of reacting to known events, the organization can develop work processes based on predictions (Agrawal et al., 2018). For example, predictive policing enables a police force to forecast where future crimes may occur (Shapiro, 2017; Waardenburg et al., 2022). This requires new processes within the police force, such as setting up preventive measures before the crime occurs, for instance, by proactively increasing the police presence in the area (Brayne, 2017; Shapiro, 2017; Waardenburg et al., 2022). AI predictions allow organizations to decide actions for future events and to expend less effort preparing for unknown probable events (Agrawal et al., 2018). Thus, integrating AI across the organization can result in changing processes and even entire operating models as new tasks become necessary to capture the benefits of predictions (Agrawal et al., 2018; Iansiti & Lakhani, 2020).

Second, AI is a general-purpose technology (Brynjolfsson & Mitchell, 2017). As such, it can be applied in all parts of the organization and for various purposes. However, AI has been called weak or narrow, as an AI application can only solve a specific task or problem (Iansiti & Lakhani, 2020). For instance, an AI application developed to play GO cannot also edit text. Thus, an organization striving to become AI-driven will not develop one AI application to solve all problems but instead set up many applications where each performs a single task in a network of other AI applications (Iansiti & Lakhani, 2020). Development depends on data access, where the output from existing AI applications can be used for improving or developing other AI applications (Iansiti & Lakhani, 2020). As such, integrating AI applications may create opportunities for new applications that also demand changes to existing tasks and processes. Third, AI has an inherent, built-in dynamic, as it can continuously learn from new data. Compared to traditional IT software, which only changes when a software update is introduced, AI can alter its output based on new data. One reason for wanting AI applications to learn from new data is to improve their accuracy or generate new insights (Grønsund & Aanestad, 2020; Iansiti & Lakhani, 2020). However, continuous learning can also deteriorate results if the data is of low quality, contains faults, or is skewed (Boyd & Crawford, 2012; Danks & London, 2017; Grønsund & Aanestad, 2020).

AI's inherent dynamic aspect also involves its output inconsistency. Continuously learning from data, AI does not necessarily display consistency in its output, which may generate surprising results (Metz, 2016). For example, during a GO match between Google's DeepMind AI, AlphaGO, and the renowned GO player Lee Sedol, the AI system surprised the entire GO community by making a move that had never been seen before. At first, the move was perceived as a mistake, but it proved an accurate calculation as AlphaGO went on to win the match (Metz, 2016). Such surprising output can generate uncertainties since assessing accuracy can be challenging (European Society of Radiology, 2019). For instance, going back to the example of predictive policing, by proactively sending out police officers to an area where a crime is expected to be committed, police presence alone may prevent a crime. However, the only evidence that the presence of police officers prevented the crime is its absence, which is the same outcome as no crime being expected in the first place. Furthermore, turning to AI and expecting it to explain or share its reasoning can prove challenging. Some AIs are black-boxed due to their complex internal logic (Castelvecchi, 2016; Rudin, 2019) or hidden as intellectual property (O'Neil, 2016).

Transforming work

As AI becomes more common in organizations, scholarly interest in how AI will transform work and organizations has increased (Berente et al., 2021; Faraj et al., 2022; von Krogh, 2018). Where some scholars explore whether AI will automate work tasks (Frey & Osborne, 2017) or augment employees at work (Rai et al., 2019; Raisch & Krakowski, 2020), others have focused on how AI will alter employees' jobs and tasks (Faraj et al., 2022; von Krogh, 2018). For instance, AI professionals can begin to handle a task previously performed by another professional (Faraj et al., 2018; Galperin, 2017). That technology can allow tasks to shift is nothing new. For instance, as the functionality of the gastrointestinal endoscopy evolved, it allowed gastroenterologists to treat pathologies, a task previously exclusive to surgeons (Zetka, 2001). However, with AI, non-professionals can now perform the task of knowledge workers. For instance, seasonal tax preparers can use an expert system to perform the work of accountants (Galperin, 2017).

Another string of research has explored the adverse effects of integrating AI in organizations, such as employers' increased control over employees (Faraj et al., 2022; Kellogg et al., 2020) and the negative implications of using AI for people analytics, such as performance evaluation and personal development (Giermindl et al., 2022). One example of an adverse effect is how employees start to perceive their abilities as inferior to AI and accept the authority of the AI or the roles that represent it (Introna, 2016; Waardenburg et al., 2022). For instance, after introducing Turnitin's text-matching algorithm, which detects plagiarism in academic writing, tutors began questioning their own expertise. Having access to the service, they began to submit all student work to the service, which they perceived as more objective and better at detecting plagiarism (Introna, 2016). Introducing AI to aid professionals can also result in resistance (Brayne & Christin, 2021; Kellogg et al., 2020). For instance, in a study of web journalists and legal professionals being presented with AI to aid their work, both groups resisted the impacts of AI by developing buffering strategies, such as foot-dragging (Christin, 2017).

As AI needs to be part of work processes, building relevant AI applications demands the involvement of both domain and technical data science expertise (Fountaine et al., 2019; Grønsund & Aanestad, 2020; Pachidi et al., 2021). Domain experts are vital because they are the ones that know the business processes and desired goals and outcomes. For instance, during the development of an AI application to support an organization's hiring process of job candidates, the developers had to include the domain experts and their knowledge in selecting data, understanding the hiring process, and ensuring that the AI application realized the vision of the workforce (van den Broek et al., 2021).

Trust in AI

Employee trust in AI is perceived as critical to successfully integrating AI in organizations (Candelon et al., 2021; Fountaine et al., 2019; Glikson & Woolley, 2020). Trust is commonly defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau et al., 1998, p. 395). Trust relations mainly involve two roles: the trustor, a trusting party, and the trust referent, the trusted party, also referred to as a trustee. Critical to trust is that the trustor must have positive expectations, held in the presence of risk or uncertainty, calling for the trustor to accept vulnerability (Fulmer & Gelfand, 2012). In terms of trust in AI, these roles are often held by a human trustor and a single AI as trust referent (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021). Furthermore, research has revealed that trust in AI often depends on AI-specific capabilities, such as its transparency and reliability, as well as the level of task characteristics (Glikson & Woolley, 2020; Lockey et al., 2021).

Transparency is crucial to trust in AI, which refers to users receiving information regarding the AI's inner logic (Glikson & Woolley, 2020; Hoff & Bashir, 2015). However, more advanced AI can suffer from AI being black-boxed and incapable of explaining itself, and even a data scientist can have difficulties revealing how the AI came to its conclusion (Castelvecchi, 2016). Two recent literature reviews highlight the importance of AI transparency for building trust in AI (Glikson & Woolley, 2020; Lockey et al., 2021). Both reviews emphasize the need to explain the AI's choices and reasoning for users to develop trust in the AI. For instance, communication regarding the AI's rationality can improve the calibration of users' expectations of AI (Glikson & Woolley, 2020). However, the downside of explanations is that they can lead to over-trust and can manipulate the user into assigning trust in the AI when trust is not warranted (Lockey et al., 2021). Nevertheless, research also shows that people may be willing to rely on AI even if it is black-boxed (Logg et al., 2019). Thus, the impact of transparency on trust for AI remains unclear.

Reliability refers to AI exhibiting consistent and expected behavior over time (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021; Lumineau et al., 2022). Important to notice is that the AI's actual reliability is not enough to warrant the user's trust; research shows that the users must also perceive it to be reliable (Lockey et al., 2021). Nevertheless, assessing AI's reliability can be challenging, as AI depends on data and can continue to learn and improve or deteriorate its accuracy over time, as mentioned earlier.

Research also reveals that trust depends on the task the AI is developed to perform, referred to as task characteristics or task substitution, including augmentation and automation (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021). Depending on the task, a trustor can have different preferences for human or AI actors. For example, research found that AI fairness and trustworthiness were attributed to AI being objective (Lee, 2018). However, when it came to tasks demanding human skills (e.g., subjectivity and emotional capabilities), such as making decisions on hiring or evaluating work performance, AI was perceived as less trustworthy (Lee, 2018).

Trust also depends on the level of autonomy the AI is given, whether it is entrusted to perform a task without human supervision, or whether there is a "human-in-the-loop" (Lockey et al., 2021). For instance, having a doctor-in-the-loop for AI in health recommendation systems can improve acceptance (Calero Valdez et al., 2016).

Distrust in AI

Distrust is a separate construct from trust, where low trust is non-equivalent to distrust (Dimoka, 2010; Lewicki et al., 1998; Saunders et al., 2014). Distrust is defined as a "confident negative expectation regarding another's conduct" (Lewicki et al., 1998, p. 439), where the trustor (or distrustor) is unwilling to succumb to vulnerability (Bijlsma-Frankema et al., 2015; Lewicki et al., 1998). To the best of my knowledge, distrust in AI is a relatively unexplored phenomenon. However, research has demonstrated how users can reject AI or develop algorithmic aversion (e.g., Christin, 2017; Dietvorst et al., 2015), which is similar to distrust, whereby a trustor is unwilling to succumb to the vulnerability of the trust referent. Rejection can occur when AI capabilities are not met. For instance, the lack of transparency can result in users attempting to resist or game the AI (Möhlmann & Zalmanson, 2017). In terms of reliability, an AI erring can result in algorithmic aversion, where a user prefers human aid over AI (Dietvorst et al., 2015). Overcoming trust breakdowns can require trust repair, such as explaining what went wrong (Dzindolet et al., 2003) or normalizing reliability breakdowns (Karunakaran, 2022). Lastly, regarding task characteristics (Glikson & Woolley, 2020), research has revealed that domain experts are more reluctant to rely on AI output than laymen (Hoff & Bashir, 2015; Logg et al., 2019). One reason for domain experts' skepticism may be that they perceive a risk of deskilling and loss of work security because AI can perform tasks independently (Lockey et al., 2021).

The challenges of bias

As noted above, the perception that AI is objective (Lee, 2018) is one of the reasons for arguing that AI is preferred for specific decision-making (Agrawal et al., 2018). Research has shown that human decision is riddled with subjectivity and bias regarding race, age, social class, or gender (Bertrand & Mullainathan, 2004; Brewer & Lui, 1989; Rivera & Tilcsik, 2016). Expecting humans to mitigate their biases is not necessarily possible or realistic, as judging others, for instance, based on age and sex, can be done unconsciously and almost instantaneously (Brewer & Lui, 1989; Ridgeway, 2006; Stangor et al., 1992). Handling bias is necessary, as untreated bias leads to problematic and undesirable outcomes. For example, in one study, students were asked to rate the assisting instructors of an online course. The study revealed that the students rated the assistants they perceived as male significantly higher than those perceived as female, demonstrating gender bias in their evaluation (MacNell et al., 2015). In a company, bias can have a negative impact on wage distribution processes (Hultin & Szulkin, 1999), hiring decisions (Petit, 2007), and evaluation of work achievements or assessments (Heilman, 2001; MacNell et al., 2015).

Understanding how bias impacts decision-making is not necessarily easy, as biases can reinforce or counteract each other. For instance, the drawback of one cognitive bias can trump the boost of other biases, such as gender and social class. In a résumé audit, researchers found that higher class origin provided an advantage to male applicants in the elite labor market. In contrast, women did not experience the same boost in the evaluation, as they faced a competing negative stereotype of being female and less committed to full-time, intensive careers (Rivera & Tilcsik, 2016). A further complexity is that human decisions are inconsistent and affected by context. For instance, in a study exploring judges' likelihood of granting parole to prisoners, results showed that judges were more likely to make favorable rulings at the beginning of the day, or after a break, than at the end of a parole decision session (Danziger et al., 2011).

AI is seen as favorable to overcoming human biases because it is not plagued by human weaknesses, such as prejudice, fatigue, or hunger (Agrawal et al., 2018; Frey & Osborne, 2017; Henke et al., 2016). However, this idea has somewhat fallen into disrepute as a growing field of research has established that AI can hold biases of its own (Danks & London, 2017; Ferrer et al., 2021). Such biases can emerge when AI is trained on data that contain existing human biases (Brayne, 2017; Danks & London, 2017). For instance, researchers demonstrated that when doing an image search for different occupations, not only did the search results replicate the gender distribution represented, such as displaying images of males for male-dominated occupations, but stereotypes were exaggerated. The search results also showed more male images for male-dominated professions compared to actual work distribution (Kay et al., 2015).

These challenges have not been solved with the current development in generative AI, such as GPT-4 and DALL-E (Maslej et al., 2023). Instead, the issue has increased in complexity as AI creates new synthetic data, such as AI-created humans which are fictive and lacks both gender and ethnicity (Luccioni et al., 2023). Bias can also emerge unintentionally as part of the AI development and selection of training data (Boyd & Crawford, 2012; Danks & London, 2017; Faraj et al., 2022). For instance, in a beauty contest marketed as entirely objective and with an AI judge, the result favored white western women. The company behind the AI beauty contest argued that the biased result may have been due to the low representation of minorities in the training data (Levin, 2016).

Being dynamic and learning from new data, AI can develop bias over time. One infamous example is Microsoft's chatbot, Tay. Soon after its launch on Twitter, Tay was targeted by a group of Twitter users, who fed it racist and offensive slurs and taught Tay to tweet the same (Faraj et al., 2018; Lee, 2016). Using AI for decision-making is problematic when AI biases are left unchecked, as they can lead to unfair and unequal treatment of individuals (Ferrer et al., 2021; O'Neil, 2016). For instance, this can occur when AI is used in hiring processes (Barnes, 2019) or in granting bank loans (Hale, 2021). Researchers claim that AI bias is more worrisome when AI is applied uncritically, or there is no "human-in-the-loop" (Danks & London, 2017; Teodorescu et al., 2021). To mitigate that AI involves a morally problematic situation, such as males being favorably treated in job recruiting, AI must be designed and tested carefully. However, there are both technical and awareness challenges to ensuring that AI training data is bias-free (Stone et al., 2016). One solution to mittigate biased results can be to deliberately introduce biased training data to the AI so that the AI outcome will reach the targeted moral standard (Danks & London, 2017).

Chapter 3

Research design

This thesis contains three studies, each with a unique dataset: a multi-national survey, a longitudinal case study, and a field experiment. This chapter explains the data collection, analysis, and choices I made regarding the research design related to the respective datasets. I close this chapter with a reflection on my role as an industrial Ph.D. student.

An overview

Paper 1 is a multi-national survey with five countries spanning three continents. The study explored the status of AI development, implementation, and use in medium-sized and large organizations. Thus, Paper 1 provides an overview of practices and challenges related to the organizational use of AI. Paper 2 draws insights from a longitudinal case study and data collection based on semi-structured interviews and real-time observations, as well as supplements the collection of internal corporate documents and corporate communications. The aim of paper 2 was to explore trust in AI within the context of an organization striving to become AI-driven. Lastly, Paper 3 reveals insights from a field experiment. The experiment was conducted at a single firm and included participants (employees) from multiple countries. The purpose of Paper 3 was to investigate bias in idea evaluation. An overview of the different papers and the data collection can be seen in Table 1.

Table 1. Overview of the data collection, empirical context, and paper status

Research item	*Paper 1: Getting AI Implementation Right: Insights on Challenges and Solutions from a Global Survey*	*Paper 2: Amongst a Multitude of Algorithms: How Distrust Transfers Between Social and Technical Trust Referents in the AI-driven Organization*	*Paper 3: Blinded by the person? Experimental evidence from idea evaluation*
Method	Multi-national survey	Case study	Field experiment
Type of data	Survey data	Observations, transcripts, documents	Experiment data
Amount/length	2525 respondents	24 months + 6 months of follow-up data	38 test subjects
Empirical context	Multiple firms, globally distributed	A single firm, Local site	A single firm, globally distributed, online
Geographical spread	China, Germany, India, the UK, the US	Data collected in two European countries	Global spread of subjects, including Sweden, the US, India, and China
Status	Accepted, California Management Review	Accepted to HICSS conference	Published, Strategic Management Journal

Paper 1: The multi-national survey

Our aim for Paper 1 was to better understand the challenges organizations encounter when implementing AI and the strategies they use to overcome them. In the survey, we asked respondents to share information about their organizations' historical and current initiatives to implement AI. In addition, survey questions addressed the challenges encountered during AI implementation, strategies to overcome those challenges, and plans for investment and hiring. The general idea and orientation for Paper 1's investigation originated from the case study described in Paper 2. Following the first set of interviews in the case study, the initial insights were summarized into topics regarding the main challenges, strategies, and general perceptions of becoming AI-driven and used as a basis for a qualitative interview guide. The insights were then confirmed as relevant by one of the top managers at the case study organization and further validated by an additional 15 semi-structured interviews with decision-makers or experts from other organizations. These experts and decision-makers were either currently implementing AI or advanced analytics in their organization or responsible for supporting other organizations implementing AI (see Table 2).

Table 2. Expert interviews for Paper 1

Role	*Domain*
Head of Research	Suppliers of AI/AA technology
Senior Management Consultant	Suppliers of AI/AA technology
Global Industry leader & Marketer	Suppliers of AI/AA technology
Industry advisor	Suppliers of AI/AA technology
CDO	Finance
CDO	Finance
Chief Product Owner (Data Lake)	Finance
Head of AI strategy	Finance
CTO	IT & Technology
Head of Media & Communication	IT & Technology
Strategic Product Management	Telecommunication
COO	Telecommunication
Head of Strategy	Insurance
Node Manager	AI innovation center
Founding partner	Fintech
Founding director	Fintech industry organization

Data collection

The findings revealed challenges and strategies in the technological, organizational, and cultural domains. Based on the interviews and the domains, we devised a questionnaire (see Appendix) targeting decision-makers in organizations currently implementing AI. Using panels of online business professionals adhering to ESOMAR[3] quality controls, we surveyed 2,525 decision-makers with experience implementing AI in five markets: China, Germany, India, the UK, and the US. The target quota of 500 decision-makers with AI implementation experience from each country was subdivided to survey at least 250 technical managers and 250 operational managers (see Table 3).

Table 3. Respondent per country for Paper 1

	China	*Germany*	*India*	*UK*	*US*	*Total*
Technical managers	250	259	250	257	250	1266
Operational managers	252	252	254	251	250	1259
Total	502	511	504	508	500	2525

Analysis

From the responses, we distinguished two subcategories of AI Experienced firms and AI Newcomers based on their self-assessment. Next, using these two groups to contrast each other, we analyzed their reported similarities and significant differences regarding experienced challenges and strategies concerning AI implementation and their plans for investment and hiring.

Paper 2: The longitudinal case study

Paper 2 studies how distrust in AI evolves during an organizational transformation to become AI-driven. To follow the development and consequences

[3] European Society for Opinion and Marketing Research. Responsible for guidelines and international standards for ethical and professional conduct for data-driven projects. Developed in collaboration with network of partners, including the ICC and Global Business Research Network (GRBN). https://esomar.org/codes-and-guidelines

over time, my co-authors and I conducted a longitudinal case for 24 months, using multiple data collection methods.

Empirical context

The case study follows a global technology firm (in the paper called GlobalTech) that sells and manages field equipment to business customers. Choosing GlobalTech as a suitable case study rested on one of the firm's business unit's initiatives to become AI-driven. The organizational transformation involved launching a new operating model and altering work processes, roles, and responsibilities. It also included the introduction of a new mindset the business unit wished employees to adopt. The new mindset comprised expected behaviors such as becoming 'data-driven,' meaning that employees should use analytics to continuously learn and adapt and to move the operational work from reactive to predictive. We were granted access to the organization's operation center in Europe, which allowed us to follow the transformation up close. During our fieldwork, we followed a local data analytics team whose assignment was to build analytics models for the organization, including AI.

Data Collection

We were granted access to GlobalTech's business unit in May 2019 and began our data collection that month. Table 4 provides an overview of the primary data collected during the fieldwork (see Table 4). Our initial setup was to conduct observations and interviews at the research site. As such, between May 2019 and January 2020, we spent 16 days on-site and arranged interviews and discussions on-site and in-between visits. During this time, we observed the work of several different teams. However, almost a year after we were granted access, the Covid-19 pandemic began to spread across Europe, and GlobalTech issued a work-from-home policy. This new policy also terminates our efforts to conduct on-site observations, with the result that all fieldwork was conducted online. Fortunately for us, during our previous stays, we had established strong connections with several key informants, making the transition to digital fieldwork unproblematic. In April 2021, approximately two years after we began our fieldwork, the organization made a local transformation, including splitting the data analytics team we were following.

Therefore, this team split formed a natural end to our fieldwork. However, we carried on our semi-structured interviews with key informants from May to December 2021 to capture their reflections on the split and the new organizations.

Table 4. Data collection for Paper 2

Year	*Type of data*	*Frequency/Amount*
2019	Days spent at the operation center	13 days
	Times spent on observations	9 days
	Observation at GlobalTech HQ (HR workshop)	2 days
	Recorded interviews and discussions	26
	Documented discussion (field notes)	9
	Documents collected	50
2020	Observations at the operation center	3 days
	Digital observations (online meetings)	5 hours
	Recorded interviews and discussions	25
	Documents collected	36
2021	Recorded interviews and discussions	18
	Documents collected	21 (+ 3 during 2022)

Analysis

For our analysis, we used the principles of constant comparisons in grounded theory (Glaser & Strauss, 1967), where we followed established guidelines for inductive concept development (Gioia et al., 2013) to ensure the rigor of our research results. We built a data structure and coded data into first-order concepts, second-order themes, and aggregated dimensions (Gioia et al., 2013). As trust emerged as a central phenomenon, we continued by following Brattström et al.'s (2019) identifying statements and behaviors that expressed either positive or negative trust perceptions (Lewicki et al., 1998). We began to compare our themes with existing concepts from the literature, such as trust in AI (Glikson & Woolley, 2020), trust development (Lewicki et al., 2006), and trust transfer (Stewart, 2003), and we looked for similarities and differences that could explain our phenomena. During this process, we noticed that our case was displaying the construct of distrust (Bijlsma-Frankema et al., 2015; Lewicki et al., 1998; Sørensen et al., 2011) rather than trust. Further distilling our second-order themes into four aggregated dimensions, we revealed three distrust dynamics that shaped trustors' views of AI during the digital transformation.

Paper 3: The field experiment

For Paper 3, we used a field experiment to investigate bias in idea evaluation. The primary study was conducted at our partner firm. The field experiment showed a null result, and an additional online experiment using Prolific was subsequently conducted.

Empirical context

We conducted the field experiment at our partner firm, inviting innovation managers to evaluate ideas other employees had proposed through the firm's idea management system. The participants were recruited from the firms' existing network of innovation managers. The network is developed to stimulate innovation within the firm and is open for applications from all employees, irrespective of their position, unit, or location. The innovation managers are part of what drives the innovation community, whose primary purpose is to evaluate ideas at the first step of the idea management system. The idea creators, who post their ideas to the idea management systems, can be any employee within the firm. The system is open to any idea without prescreening for quality or topic. Access to this empirical setting for conducting our experiment allowed us to investigate the presence of bias in idea evaluation within an organizational setting.

Research design

The innovation managers evaluated ideas under two conditions: blind and non-blind evaluations. In the blind evaluation, the innovation manager received no information about the idea proposer. Instead, they saw "Submitted by: N/A." The innovation manager receives information about the idea proposer in the non-blind evaluation. This information included name, organizational unit, and geographical location. We used a within-subject design in which each innovation manager evaluated ideas under both conditions. We collected 412 ideas from the idea management system. Most of the ideas were categorized under four headlines: autonomous vehicles (124 ideas), design thinking (87 ideas), logistics (86 ideas), and smart manufacturing (64 ideas). All participants were asked to evaluate 48 randomly assigned ideas from 412 available ideas. In total, 60 innovation managers were recruited; of these, 38 completed the evaluation of all 48 ideas, and eight evaluators began the idea evaluation but did not complete it. In addition, the participating innovation managers answered an exit survey following the online field experiment. Findings from this survey are not presented in the paper but are included in this thesis.

Analysis

As a first step, we conducted mean comparisons between the treatment conditions where the difference was small (0.0636), and t-tests failed to reject that blind and non-blind evaluations produce the same mean outcomes. The distributions of overall scores exhibit no clear differences between the blind and non-blind conditions, either in the middle or in the tails. As this null result surprised us, we conducted an additional online experiment where we replicated the field experiment as closely as possible and took the same steps to analyze the data as in the field experiment. In neither design did we find differences in the evaluation scores of ideas proposed by women or men compared to those in the blind condition.

The position of an industrial Ph.D. student

For any researcher, pre-understanding one's position is vital for identifying how it influences the research (Alvesson & Sandberg, 2013). As an industrial Ph.D. student, this includes reflecting on the two positions this role includes: practitioner and academic student. In a sense, the position of the industrial Ph.D. student resembles a boundary-spanning role (Aldrich & Herker, 1977;

Levina & Vaast, 2005; Pawlowski & Robey, 2004). According to Barley (1996), "technicians stood with one foot in the material world, and the other in a world of representations" (Barley, 1996 p. 418); an industrial Ph.D. student could be said to stand with one foot amongst the practitioners and the other finding its foothold in academia. This dual position comes with both advantages and challenges. I describe below how I have handled both.

Awareness

Being familiar with the organization where research is conducted lowers the difficulty of understanding the structure and culture of that organization. Much of what otherwise might be tacit knowledge is already known to the Ph.D. student, such as perception of management, cultural beliefs, and general internal language, such as abbreviations. Such knowledge is helpful in observations and interviews. However, there is a challenge with being already part of something, as it can blind you to certain things. I work for a multinational firm that is present in over 180 countries and has over 100,000 employees. However, the units granting me access to conduct my research were far from my own, not only geographically but also in terms of operational work. As such, I had the advantage of exploring organizations whose structure and culture I understood but whose work tasks and processes I was still unfamiliar with. To further mitigate missing vital information, given my proximity to the organization, I engaged in an ongoing and transparent dialog with my supervisors and co-authors to examine and question the collected data.

Assumptions

Even if the organization being studied is not entirely familiar, assumptions may still exist (Alvesson & Sandberg, 2013). As an industrial Ph.D. student with a background amongst practitioners, assumptions regarding the industry can, for instance, include the perception that a phenomenon is wholly unique (when it is not) or that a phenomenon is common knowledge (when it is not). It can also include assumptions regarding empirical data, such as the prevailing corporate narrative woven into the corporate reality fabric; for instance, what AI entails and what it means to become AI-driven. I handled this challenge in two ways: first, by having the same transparent dialog with my supervisors and co-authors, as mentioned above, and second, by developing an explicit data structure, especially in the case study, where I could compare data with the existing academic literature.

Access to data

An industrial Ph.D. student can enjoy the advantage of having access to data. For my part, not only has my employment granted me access to organizations to study and informants to interview, but it has also enabled me to create additional datasets, such as the multi-national survey for Paper 1.

Expectations

Lastly, a challenge worth mentioning as an industrial Ph.D. student is the constant need to juggle interests and expectations from both academia and one's employer. Being part of two organizations can result in conflicting ideas regarding what should be done. For instance, there may be conflicting ideas on what research questions are interesting to pursue or different ideas about reasonable time plans. I handled this by being as transparent as possible and by translating the different needs of both sides.

Chapter 4

Overview of papers

In this chapter, I provide an overview of the three papers included in my thesis. I present the abstract from each paper and elaborate on these papers' main findings. In addition, I also include additional data from studies conducted for Paper 1 and Paper 3 that were not included in the papers but which I found relevant to the overall research question for this thesis. I first present Paper 1, which is an overview of the practices and challenges related to the organizational implementation of AI, thus providing a back-drop for subsequent papers, particularly Paper 2. Paper 2 discusses the formation of distrust in AI while integrating AI into the organization. Lastly, Paper 3 explores biases in idea evaluation. Its null finding is essential to the discussion on mitigating AI bias. Furthermore, I include data from the exit survey from Paper 3, which was included for the participants in the online field experiment, as a link to the discussion of AI expectations. The data are presented in this chapter.

Paper 1: Getting AI implementation right: Insights on challenges and solutions from a global survey

Paper 1 draws insights from a multi-national survey targeting decision-makers in organizations currently implementing AI. Below I share the main findings and contributions from the study. This also includes additional data, where the decision-makers have answered questions about how they perceive

AI's transformational power. For the current status of the paper, see Table 5. This study also resulted in an industrial report targeting practitioners as an audience.[4]

Table 5. Status of Paper 1

Status of Paper 1	
Full title	Getting AI implementation right: Insights on challenges and solutions from a global survey
Authors	Rebecka C. Ångström, Michael Björn, Linus Dahlander, Magnus Mähring, Martin W. Wallin
Journal	*California Management Review*
Status	Accepted

Paper 1 abstract

The promise of artificial intelligence (AI) is pervasive, yet companies experience many implementation challenges. We surveyed 2,525 decision-makers with AI experience in China, Germany, India, the UK, and the US and interviewed 16 AI implementation experts to understand the challenges companies face when implementing AI. Our study covers technological, organizational, and cultural factors and identifies key challenges and solutions for AI implementation. We develop a diagnostic framework to help executives navigate AI challenges as companies gain momentum, manage organization-wide complexities, and curate a network of partners, algorithms, and data sources to create value through AI.

Findings

The aim of Paper 1 was to explore the concrete challenges and strategies that organizations encounter and employ when implementing AI. The idea came from findings early in the case study (Paper 2), which triggered our interest in seeing how common they were. For instance, in the case study, a recurring

[4] The report can be found online: https://www.ericsson.com/4ab2b3/assets/local/reports-papers/industrylab/doc/adopting-ai-report.pdf

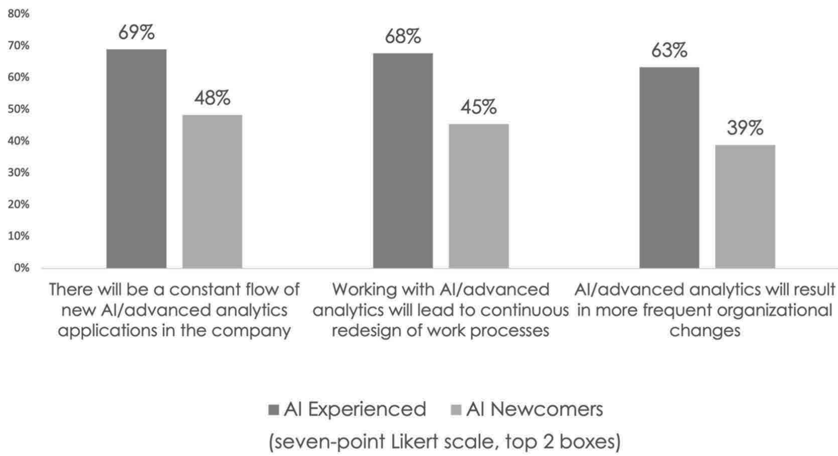
argument as to why the uptake of AI was so slow was that older generations do not understand the technology. Conducting the expert interviews as a pre-study to the multi-national survey, we picked up on the same objection from some experts when asking about the challenges of implementing AI. Adding this item to the survey, we saw that this was a common belief amongst especially AI newcomers.

Reviewing the results from the survey, we found that almost all organizations, 91 percent, have experienced challenges in all three domains that we explored: technology, organization, and culture. Some of the most frequently reported technological challenges are related to data; for instance, that data are unavailable or not structured for algorithmic use. Regarding organizational challenges, the main challenge was the lack of trained employees. Lastly, the main cultural challenges were fears and concerns amongst employees, such as fear of job loss, expertise, and autonomy. Although perceived challenges differed somewhat between the "AI Experienced" and the "AI Newcomers," the level of fear and worries among employees was the same in both groups. In terms of strategies to overcome these challenges, strategies to meet technology challenges focused on improving the usability of AI tools were most often mentioned. This was followed by improved access to data, enhanced quality of output from algorithms, and improved data management. Organizational strategies focused on organizing and governing AI innovation activities, and cultural strategies focused on skills development and change management, getting the employees to accept the transformation.

Additional data from the survey

The decision-makers also received questions regarding how they perceive the transformational power of AI (see Figure 1). Amongst the AI Experienced, almost 69 percent believe there will be a constant flow of AI applications in the company. In addition, as many as 68 percent believe that working with AI will lead to a continuous redesign of work processes. Almost as many, 63 percent, believe that implementing AI will result in frequent organizational changes. Among the AI Newcomers, the perception of the transformational power of AI is not as pervasive, as just above two out of five express the same beliefs.

Figure 1. Expectations on AI transformation



Paper 2: Amongst a multitude of algorithms: How distrust transfers between social and technical trust referents in the AI-driven organization

Paper 2 draws insights from a longitudinal case study conducted at a firm transforming to become AI-driven. To be consistent with Paper 2, I retain the term 'algorithm' for the specific applications used in the organization. However, the definition accords with that of AI in this paper. For the current status of the paper, see Table 6.

Table 6. Status of Paper 2

Status of Paper 2	
Full title	Amongst a multitude of algorithms: How Distrust Transfers Between Social and Technical Trust Referents in the AI-driven Organization
Authors	Rebecka C. Ångström, Magnus Mähring, Martin W. Wallin, Eivor Oborn, Michael Barrett
Target Journal	TBD
Status	A shorter version of this paper is accepted to the Hawaii International Conference on System Science (HICSS) -57

Paper 2 abstract

Although trust has been identified as critical for successfully integrating artificial intelligence (AI) into organizations, we know little about trust in AI within the organizational context and even less about distrust. In this paper, we investigate how distrust in AI unfolds in the organizational setting. We draw from a longitudinal case study in which we follow a data analytics team assigned to develop numerous AI algorithms for an organization striving to become AI-driven. Using the principles of grounded theory, our research reveals that different organizational distrust dynamics shape distrust in AI. Thus, we develop three significant insights. First, we reveal that distrust in AI is situated and involves both social and technical trust referents. Second, we show that when a trust referent is rendered partly invisible to the trustor, this leads to the misattribution of distrust. Lastly, we show how distrust is transferred between social and technical trust referents. We contribute to the growing literature on integrating AI in organizations by articulating a broader and richer understanding of distrust in AI. We present a model of distrust transference actuated by social and technical trust referents. We also contribute to the literature on trust, showing how AI artifacts are implicated in trust relations within organizations.

Findings

Following an organization as it develops algorithms to aid its operational work, we identified three distrust dynamics. In Distrust dynamic 1, we found that operational employees develop distrust in the corporate AI strategy as the organization stresses AI capabilities. At the same time, the impact of AI and its practical implications are unknown. The distrust generated challenges for the data analytics team as they introduced their algorithms. In Distrust dynamic 2, operational employees who engaged in developing algorithms recognized that the data analytics team lacked domain expertise, while their technical expertise was invisible. The unbalanced view of their expertise generated distrust in their ability to build relevant applications. Lastly, in Distrust dynamic 3, the operational employees are unaware of the algorithms' dependency on data (data quality or data access) and develop distrust towards the algorithms when they do not fully understand the algorithms' limitations. The distrust towards the developers and the algorithms transferred between the two, creating a cycle of distrust. Lastly, in our case study, we noticed that both the data analytics team and the algorithms are partly invisible to operational employees, whereas data as trust referent is entirely invisible, resulting in misattribution of distrust.

From the case epilogue

The description of the case in Paper 2 ends with the Covid-19 pandemic. The pandemic became an exogenous shock when the firm issued new directives forcing employees to work from home. Hence, the distrust dynamics also evolve. Removed from the physical environment, the organization found itself dependent on data analytics to run the operational work. Thus, the operational teams started requesting the data analytics team to develop algorithms, and their increased experience with these algorithms built the employees' trust in them. However, with expertise came new expectations from algorithms, from addressing internal efficiency to establishing new customer business values. As a result, there were new trust barriers to overcome as more algorithms were being developed.

Paper 3: Blinded by the person? Experimental evidence from idea evaluation

Paper 3 presents the findings from a field experiment conducted at a global tech firm. Below I share the main findings and contributions from the study.

I also share data from the exit survey that targeted the participants of the field experiment. For the current status of the paper, see Table 7.

Table 7. Status of Paper 3

Status of Paper 3	
Full title	Blinded by the Person? Experimental Evidence from Idea Evaluation
Authors	Linus Dahlander, Arne Thomas, Martin W. Wallin, Rebecka C. Ångström
Journal	*Strategic Management Journal*
Status	Published

Paper 3 abstract

Seeking causal evidence on biases in idea evaluation, we conducted a field experiment in a large multi-national company with two conditions: (a) blind evaluation, in which managers received no proposer information, and (b) non-blind evaluation, in which they received the proposer's name, unit, and location. To our surprise—and in contrast to the pre-registered hypotheses—we found no biases against women and proposers from different units and locations, which blinding could ameliorate. Addressing challenges that remained intractable in the field experiment, we conducted an online experiment, which replicated the null findings. A final vignette study showed that people overestimated the magnitude of the biases. The studies suggest that idea evaluation can be less prone to biases than previously assumed and that evaluators separate ideas from proposers.

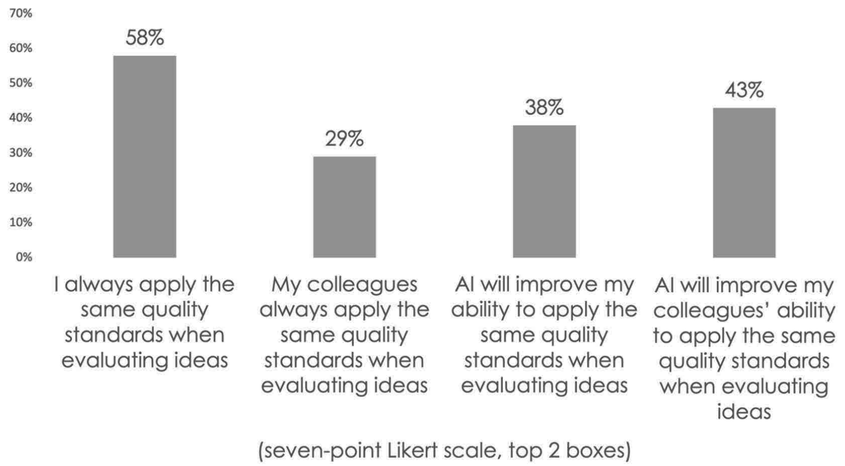
Findings

Our field experiment at our partner firm generated a null finding where we found no biases against women and proposers from different units and locations during the idea evaluation. We elaborated on four plausible reasons for this result. a) That the findings resulted from the organizational culture achieved at our partner firm. However, as a result, it was replicated in the online experiment, and the null finding persisted, contradicting the absence of bias being connected to the organization's culture. b) Selection into the experiment could lead to a null finding, and people who are positively inclined to give women and members of the outgroup higher evaluations would have been selected to be part of our experiment. However, the firm identified the group of innovation managers participating in our experiment, and there was no opportunity to select for the online experiment based on being more lenient towards disadvantaged groups. c) Separation of idea from person. Our null finding may have arisen because the idea takes precedence over the person; the proposer's identity does not evoke information critical to idea evaluation. Even when we made the proposer's gender more salient in the online experiment, we found no gender differences. d) Shifting standards from evaluation to selection where there is a difference between evaluation and selection. For instance, when evaluating job candidates, a female candidate may be seen as "good for a woman." We studied the evaluation of early-stage ideas, which still have a long way to go before eventual selection. The evaluators did not make the final selection and had few budget constraints, which may reduce biases. This could explain why our results differed from previous work focused on selection.

Additional data from the exit survey

Paper 3 is an inquiry into human bias (or lack of bias), a field of knowledge critical for building trustworthy AI. All participants completing the experiment were presented with an exit survey investigating the innovation managers' view of AI. This survey reveals that most idea evaluators are confident that they can apply the same quality standards when evaluating ideas (see Figure 2). However, fewer have the same confidence in their colleagues. Furthermore, almost 38 percent of evaluators believed that their ability to apply the same quality standards would improve with the help of AI. Forty-three percent also believed that their colleagues' abilities would improve by being helped by AI.

Figure 2. AI and quality standards in idea evaluation



Chapter 5

Key findings and theoretical implications

As organizations strive to become AI-driven, we will all need to learn how to dance with the machine. What makes this dance different from previous technology integrations is the dynamics that AI embodies, which catalyzes transformative changes in the organization. These dynamics demand that we become even more flexible and ready to adapt to the alteration of our work. However, if the dynamics are poorly understood, the result can be distrust, resistance, and an unwillingness to engage in this dance. In the following discussion, I summarize the key findings from my research and their implications for future research.

Summary of key findings

AI is advancing its position in organizations, performing tasks previously exclusive to humans. Following this development, scholars have explored how the introduction of AI will alter work for decision-making, expertise, and identity, to name only a few areas (Faraj et al., 2018; Vaast & Pinson-neault, 2021; von Krogh, 2018). Scholars have also identified trust as critical for successfully integrating AI into the organizational context (Glikson & Woolley, 2020; Leonardi et al., 2022; Lockey et al., 2021). Without establishing trust, employees may develop strategies to resist AI integration (Brayne & Christin, 2021; Christin, 2017; Kellogg, 2020).

What is unique about AI, as stated in the introduction to this thesis, is its embodiment of three different dynamics that drive transformational change in organizations. First, organizations must adapt and adjust work tasks and processes to proactively act on predictions (Agrawal et al., 2018). Second, AI is a general-purpose technology and can be applied across organizations for various purposes. As such, an organization wanting to become AI-driven will likely develop numerous AI applications across the organization (Iansiti & Lakhani, 2020). Third, AI has an inherent dynamic allowing it to learn from data and change its output over time (Faraj et al., 2018). Taking this perspective and drawing from the three papers in my thesis, I offer three insights below into how AI's dynamics challenge trust and may halt AI integration. First, becoming AI-driven is a continuous transformation where challenges persist over time, including employee distrust towards AI. Second, the continuous development of numerous AI applications results in an ongoing interplay between social and technical trust referents, which can result in vicious distrust cycles. Third, the lack of understanding of AI's inherent dynamic—continuously learning from data—can result in unrealistic expectations of AI and lead to distrust.

A continuous transformation

As a general-purpose technology (Brynjolfsson & McAfee, 2017), AI can be applied for various organizational purposes (Agrawal et al., 2018). Thus, by becoming AI-driven, an organization is expected to develop numerous AI applications across the organization (Iansiti & Lakhani, 2020). However, this insight leads one to ask: When will the AI transformation be complete? Finding an answer is challenging. According to the findings in Paper 1, a large majority of 'AI Experienced' firms foresee a constant flow of algorithms being developed. They also foresee that introducing these algorithms will shift work processes and trigger organizational transformations in the foreseeable future. The 'AI Newcomers,' on the other hand, are not as convinced that AI will lead to such a continuous transformation in their organizations. The difference indicates that insights into organizational impact come with experience. For instance, organizations learn from experience that obtaining access to new data, or AI applications generating new insights, can result in more opportunities to develop AI.

As work processes shift, employees' work tasks become altered. It is, in a sense, similar to what Kevin Kelly, former executive editor of WIRED magazine, describes in his book The Inevitable as a cycle of "robot replacement," where robots or computers gradually overtake humans' tasks and, eventually, jobs. In the process, humans invent new things they wish to do, which again become gradually overtaken by robots or computers as their abilities evolve (Kelly, 2017). Research shows that organizational disruptions, for instance, triggered by technological advancement, can result in employee distrust in the organization (Gustafsson et al., 2021; Sørensen et al., 2011). Research also shows that AI's impact on work is a key factor for trust in AI (Gillespie et al., 2021) and that introducing AI into the workplace can be met with resistance (e.g., Brayne & Christin, 2021). Thus, we know there might be initial challenges to establishing trust when integrating AI into the organizational context. However, research does not explain how trust or distrust evolves following a continuous AI transformation. Nevertheless, Paper 1 informs us that employees in both 'AI Experienced' and 'AI Newcomers' firms experience fear of loss of jobs, expertise, and control. This similarity indicates that the challenge of establishing trust persists over time. Rather than decreasing as organizations continue to develop and integrate AI, the fear of losing jobs, expertise, and control continues to challenge these firms.

This insight also points to the importance of conducting longitudinal studies of trust in AI, especially as research exploring how trust and distrust in AI develops over time is scarce and often focuses on a single AI application (Glikson & Woolley, 2020).

Continuous development enables distrust cycles

Beyond the impact of work, scholars studying trust in AI point out that specific capabilities, such as reliability, transparency, and task characteristics, are vital for establishing trust in AI (Glikson & Woolley, 2020; Lockey et al., 2021). At the same time, failure to demonstrate these capabilities can lead to resistance. For instance, research has shown that perceiving the AI to err can lead to algorithmic aversion (Dietvorst et al., 2015), and the lack of transparency can result in users attempting to resist, or game, the application (Möhlmann & Zalmanson, 2017). However, few scholars, if any, have explored trust or distrust towards AI within organizations that develop numerous AI applications or, as previously stated, are undergoing a continuous AI transformation. Furthermore, only a few (e.g., Chawla, 2020; Leonardi et al., 2022; Lumineau et al., 2022) have focused on relations beyond the individual relationship between human trustor and the single AI trust referent (Jacovi et al., 2021; Lockey et al., 2021). This is, however, the phenomenon my co-authors and I explore in Paper 2. We show that as the organization develops numerous AI applications, social and technical trust referents are continuously involved in developing and integrating the AI applications into the organizational context. We also show that distrust transfers between these social and technical trust referents, resulting in the emergence of pervasive distrust cycles (Bijlsma-Frankema et al., 2015). In other words, the experience of developing and integrating one AI application influences the employee's perception of AI developers' ability to build further AI applications. If the experience is bad, distrust towards the developers emerges, which transfers to the AI applications they develop. Likewise, the experience of an AI application breaking down builds distrust in the application, which transfers to the developers and their future work. Thus, developing AI applications is a socio-technical (Mumford, 2006) process that does not occur in isolation. Instead, the experience of developing and integrating numerous AI applications reproduces distrust that influences future applications. As such, the second AI dynamic, stimulating the development and integration of numerous AI applications (Agrawal et al., 2018; Iansiti & Lakhani, 2020), can lead to continuous distrust formation. This is an essential insight for any scholar wishing to explore trust in AI, as it exposes the need to include context and study the phenomenon over time. One interesting path to explore would be to see if trust, like distrust, can be reproduced in a similar fashion.

Paper 2 also identifies that these AI-related distrust cycles are closely related to the domain and technical expertise levels. Research has shown that technical and domain expertise is needed in AI development (Lou & Wu, 2021; van den Broek et al., 2021). However, in Paper 2, my co-authors and I show that as the domain experts lack technical expertise, they risk misjudging both the developer's ability and the AI's ability and reliability. Misjudging the developers and the AI can further lead to misattribution of distrust. For instance, when the domain experts do not understand the technical constraints that bind the developers, such as the need for data access to build AI applications, the domain experts risk blaming the developers for doing a poor job when AI ability is hampered due to data scarcity. Misjudging the developer's ability and the AI's ability and reliability can further fuel distrust cycles if the misconceptions persist over time. Thus, fostering technical expertise, especially data and AI literacy, is vital for building trust and avoiding distrust in AI. Moreover, understanding how domain experts develop technical expertise is an important area for further exploration, which I develop below under "implications for future research."

The missed inherent dynamic

Not being familiar with AI's dependency on data requires that we learn more about data and ourselves as humans. In Paper 3, the field experiment, two crucial insights add to the discourse on trust in AI, especially regarding AI reliability. The first draws on the paper's main finding, namely the null result. Research has proven that humans are subjective and hold prejudices and biases toward one another (Bertrand & Mullainathan, 2004; Brewer & Lui, 1989; Rivera & Tilcsik, 2016). AI is said to be objective and consistent in helping us in decision-making (Lee, 2018). However, recent research has demonstrated that, as AI is trained on data often based on previous decisions, it can both inherit our human biases, incorporate them into its results, and reinforce bias in organizations (Boyd & Crawford, 2012; Brayne & Christin, 2021; O'Neil, 2016). One way to handle this is to curate data, or provide AI with skewed data, to balance out the existing biases (Danks & London, 2017). However, our field experiment investigating bias in idea evaluation showed that we may not fully understand where and how biases exist. In contrast to our expectations, we found no proof of bias towards female idea-givers in our two experiments. This raises the critical notion that if we are unaware of where and how biases influence our decisions, we cannot build tools, such as AI, to mitigate them. Our study does not deny the existence of bias amongst human decision-makers but instead illuminates the fact that we simply do not know enough. As such, any researchers exploring trustworthy AI, focusing on bias in data, must thoroughly investigate if and in what form biases exist prior to studying AI's impact.

The second insight from Paper 3 comes from the additional exit survey that we did not include in the paper. Though the survey was only answered fully by 34 respondents and should be seen more as an indication of a potential issue rather than a proven result, it is an important issue worth mentioning. Of the 34 respondents, more than one out of three (38 percent) believed AI would improve their ability to provide more consistent judgment. Moreover, 43 percent believed AI would help their colleagues' ability to apply the same quality standards when evaluating ideas. In other words, they would perceive themselves and their colleagues as more trustworthy if an AI helped them. However, their statement exposed that a substantial share of them perceived AI to provide consistency. Believing that AI provides consistency indicates a risk that the awareness of AI's inherent dynamic, being able to improve or deteriorate its output over time, is low amongst non-AI-experts. Instead, expecting AI to provide consistency reveals that AI is perceived more like traditional IT software. This lack of understanding of AI's inherent dynamic was also apparent in Paper 2, the longitudinal case study, where domain experts witnessed an AI breakdown and blamed it on the AI instead of recognizing that low-quality data from their organization caused the issue. These two insights—not being fully aware of how and where bias is present in either data or a real-life context and the lack of understanding of the inherent dynamic—make it challenging to accurately judge AI output and reliability. Moreover, as we saw in Paper 2, this can lead to misattribution of distrust.

To summarize the key findings, we know that the integration of AI in organizations will be far-reaching. Numerous AI applications will be integrated across organizations for various purposes (Berente et al., 2021; Faraj et al., 2018; Rai et al., 2019). However, while organizations need to learn how to orchestrate ensembles with human and machine workers (Recker et al., 2023), employees need to become willing to dance with the AI. Such willingness depends on trust in the dancing partner and the general understanding of the dance. Drawing from my three papers, we see that AI dynamics contribute to the formation of distrust in AI. The two first dynamics—reorganizing work processes and tasks around predictions and the possibility of utilizing AI across the organization—set off an endless transformation that generates uncertainty among employees about what work may look like tomorrow. The uncertainty results in fear of job security and loss of control, which leads to rejection and distrust of AI. AI's inherent dynamic, being able to learn and adapt from data, can also lead to distrust, as having a limited understanding of AI dependency on data creates false expectations. For instance, the expectation that AI can provide consistency or not deteriorate over time. When AI applications do not behave as expected, distrust is misattributed to developers or the AI applications, even when the fault may be sourced to the data or the perception of what the data contains. Furthermore, as relations between users, developers, and AI applications continue over time, distrust can include future applications too. To further understand how dynamics and trust and distrust evolve, we must continue building our knowledge of AI in the organizational context and over time.

Implications for future research

How we come to co-exist with AI will be interesting to follow. I picture our future dance as feedback loops where human employees create data, which is used for training AI, leading to new AI output that provides the organization with new insights, shaping the behavior of human employees, who, in turn, create new data. In this sense, we become AI co-creators as we construct new data, and AI mirrors our behavior—a dance during which we take our first stumbling steps today. However, my thesis is not an exhaustive exploration of how this dance will evolve or how AI will influence organizations. The research in this thesis only sheds partial light on the transformational power of AI and reveals only some of the implications for trust and distrust in AI. Nevertheless, it raises questions regarding our future relationship with machines and, hopefully, some ideas for future research. Below I will elaborate on three ideas beyond trust and distrust that I formulated as a Ph.D. student. I believe all three are interesting enough to investigate as part of the perspective that introducing AI is a continuous transformation. One of my favorite papers that I read during my Ph.D. studies is Davis's "That's interesting!" (1971), where Davis uses the elegant formula "what seems to be X is in reality non-X." Hence, I will use that formula to describe my three research ideas.

It is not a data lake; it is a river delta

My first research idea is purely based on a phenomenon I uncovered during my fieldwork for Paper 2, the longitudinal case study. During the fieldwork, the organization faced many challenges while becoming what they described as data-driven. One of these challenges involved the access and structuring of data. During the fieldwork, the organization was working on setting up a new single data repository: a data lake. A data lake is a "scalable storage and analysis system for data of any type, retained in their native format and used mainly by data specialists (statisticians, data scientists or analysts) for knowledge extraction" (Sawadogo & Darmont, 2021, p. 100). The argument for having a single repository for data, such as a data lake, is to overcome the challenge organizations often face with data, such as data being fragmented, incomplete, and siloed (Iansiti & Lakhani, 2020). In Paper 1, the survey data, we saw that not having structured data was among the most common technical challenges for both AI Experienced and Newcomers.

The metaphor of a data lake evokes a vast pool where data is kept and just waiting to be used. However, from my research into data usage in organizations, I believe this metaphor is deceptive and can result in misconceptions regarding AI development. As described in Paper 2, the data analytics team needed to access data from various sources. With the introduction of a data lake, the team believed they could access this data from a single source rather than scouting the organization for multiple sources. The data would be poured right into the lake. However, unstable original data sources are a challenge a data lake cannot surmount. In one event, we witnessed how one data source comprised daily emails with Excel files from a customer system. The structure of these files was not formalized; instead, their content shifted from day to day. Another example from the case study, not included in the paper, was how data scientists accessed weather forecast data from a third-party website. However, as the data scientist downloaded vast amounts of data numerous times daily, the third-party site eventually blocked her account.

Hence, even if the data streams end up in a single repository, they are not reliable or constant. We witnessed the same kind of unreliable stream flow from the data lake. New challenges emerged as the organization finally got the data lake in place. For instance, during this time, the organization was reorganized to improve the development and innovation of new AI applications. However, this reorganization was designed with roles and expertise in mind, for instance, mixing both technical and domain expertise. The management did not consider the data or data management when designing the new organization, resulting in data extraction from the lake becoming obfuscated between the different teams. As one team created a new data stream for an AI application they developed, another could create a similar stream for a different purpose, resulting in a confusing stream of streams, data ownership, and quality control.

In exploring the establishment of an AI-driven organization, the data resource is better compared to a river delta, where numerous data streams constantly arise, colliding and drying out. This flowing data challenges the organization's AI development (innovation) and can become onerous in structuring the organization. By foregrounding data, we not only cast a light on how data shape digital innovation but also on how failing to recognize the dynamic characteristics (flow) of data causes organizations to stumble when organizing for innovation. Exploring this phenomenon could contribute to the IS literature, digital innovation, and digital transformation.

It is not an exogenous shock; it is a continuous transformation

My second idea builds on Barley's (1986) nominal paper on the introduction of the CT scanner that became an exogenous shock challenging the role of and role relations in radiology departments and resulted in an alteration of organizational structures. Following Barley, scholars have continued to explore how the introduction of technology influences occupational roles, role relations, and work structure (e.g., Barley, 2015; Barrett et al., 2012; Beane, 2019, 2023; Beane & Orlikowski, 2015; Leonardi & Barley, 2010). Research has often shown that the introduction of technology has forced occupational roles to find new ways to conduct their work and identify jurisdictional boundaries. For instance, introducing robotic surgery altered the relationship between surgeons and medical trainees, forcing trainees to engage in norm- and policy-challenging practices to learn robotic surgery (Beane, 2019). Similarly, librarians came to redefine their occupational identity due to internet searches (Nelson & Irwin, 2014). Lastly, when gastrointestinal endoscopy enabled the treatment of pathologies, it gave rise to a jurisdictional conflict between surgeons and gastroenterologists, as the latter could now perform the activities perceived as exclusive to surgeons (Zetka, 2001). Typical of research on technology and occupational roles is that introducing a technology, or its new functionality, is treated as a single and exogenous event or shock (Barley, 1986). This shock triggers a phase of change whereby institutionalized occupational roles, role relations, and work structure are reshaped and finally settle in a new phase of stability.

However, as the introduction of AI is a continuous transformation, its impact on role and role relations will continue to deliver both external and internal shocks. External shocks occur when AI technology reaches breakthroughs, such as ChatGPT (Maslej et al., 2023). Internal shocks would occur due to the dynamic aspects of AI; for instance, when an organization finds new ways to implement AI, the development of numerous AI applications for vast areas, and the AI inherent dynamic. Several studies describe how an AI application, performing a task central to the overall operational work, has changed organizational work processes (Christin, 2017; Sachs, 2020; Waardenburg et al., 2022). To the best of my knowledge, with only a few exceptions (e.g., Vaast & Pinsonneault, 2021), little research has explored how a continuous transformation impacts roles and role relations. Nevertheless, the continuous development of AI, including external and internal shocks, will likely give rise to shifts, such as jurisdictional claims, that would be interesting to follow.

It is not a loss of expertise; it is a shift in expertise

Lastly, my third research idea also builds on an empirical finding from my first two papers and connects to the ongoing discourse on whether AI will automate or augment employees (Frey & Osborne, 2017; Raisch & Krakowski, 2020) and what will happen to domain expertise following AI advances. Regarding AI and trust, research has revealed that domain experts are more reluctant to use AI because they fear losing expertise (Hoff & Bashir, 2015; Lockey et al., 2021; Logg et al., 2019). Previous research has shown the retention of skills following automation; for instance, among pilots relying on cockpit automation (Casner et al., 2014). In Paper 1, the multi-national survey, both AI Experienced and Newcomers identified that employees were afraid that their expertise would be ignored or made redundant following the introduction of AI. However, in Paper 2, the case study, we saw that there is a need for domain experts to advance their technical expertise in order to make accurate judgments regarding AI. Other researchers have pointed in the same direction, showing that domain experts may need a combination of technological skills (Lou & Wu, 2021). For instance, art experts working with an art similarity-matching AI had to learn how the AI model functioned to manage when the AI's output did not align with the expert's expectation (Sachs, 2020).

That automation demands new expertise is nothing new. With the help of IT, we have been digitalizing and automating tasks, tools, and processes for some time, which has already altered domain experts' need for new knowledge. To make sense of the new information interface, employees must understand work at an abstract level and be able to build a theoretical understanding of data (Zuboff, 1985). The question I find interesting is how domain expertise must evolve with continuous transformation, given that tasks and processes may change continuously. For instance, as new AI models are introduced, challenging existing work, holding domain knowledge, technological understanding, and cognitive skills, such as the ability for abstraction, allows domain experts to set the direction for AI development within the organization. Rather than remain locked into assigned tasks and processes that may become obsolete, domain experts might assume intermediary roles among the technology, technology experts, and the operational environment.

Furthermore, by lifting domain knowledge to a more abstract level and increasing their knowledge of the technology, domain experts could move more freely among different processes in the organization as they change. In a sense, they might become experts in how to apply and manage AI in the organization. Their role would be similar to that of a manager. However, instead of managing employees with different domain knowledge, as the managers do, the previous domain experts become managers of different AI applications. This shift may protect the occupation from becoming superseded by the introduction of AI. Paradoxically, however, this will also make occupations in similar operational domains more alike and interchangeable. Although I may be taking this idea to its extreme, the shift in domain expertise would be an interesting path to explore nonetheless, preferably in a field where we see evidence of a faster continuous transformation, such as media (following ChatGPT) or the finance business.

Concluding remarks

Much has happened since 2015, when my colleague and I wrote about how AI will end the screen age. Today, with the introduction of powerful generative AIs such as ChatGPT, I would guess that most recognize the potential value of AI for work and other tasks. Still, one of my takeaways from the years spent exploring AI integration is that beyond the current AI hype, the integration into organizations has been slower than anticipated. Indeed, integrating AI is hard work, from identifying the right uses-cases and accessing qualitative data to finding common ground where employees (domain experts) trust the technology and accept it as part of the new AI-driven organization. Some of these challenges will be resolved as our knowledge and understanding of AI increase. This includes learning about AI's dependency on data and ceasing to perceive data as something abstract occurring in the background.

Nevertheless, I am sure that whatever happens next in AI will be at least as thrilling to follow. As such, I hope to continue to explore the role of AI and data in organizations. For you, dear reader, if you have read this far, I congratulate you for your perseverance and hope you gained some food for thought for your efforts.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. Boston, Massachusetts: *Harvard Business Review Press*.

Aldrich, H., & Herker, D. (1977). Boundary spanning roles and organization structure. *Academy of Management Review*, 2(2): 217–230.

Alvesson, M., & Sandberg, J. (2013). *Constructing research questions: Doing interesting research.* Sage.

Baker, F. (2018). The technology that could end traffic jams. BBC Future. https://www.bbc.com/future/article/20181212-can-artificial-intelligence-end-traffic-jams.

Barley, S. R. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly*, 31(1): 78–108.

Barley, S. R. (1996). Technicians in the workplace: Ethnographic evidence for bringing work into organizational studies. *Administrative Science Quarterly*, 41(3): 404-441.

Barley, S. R. (2015). Why the internet makes buying a car less loathsome: ow Technologies change role relations. *Academy of Management Discoveries*, 1(1), 5–35.

Barnes, P. (2019), November 10. Artificial Intelligence Poses New Threat to Equal Employment Opportunity. *Forbes*.

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring Boundary Relations: Robotic Innovations in Pharmacy Work. *Organization Science*, 23(5), 1448–1466.

Beane, M. (2019). Shadow Learning: Building Robotic Surgical Skill When Approved Means Fail. *Administrative Science Quarterly*, 64(1), 87–123.

Beane, M. (2023). Resourcing a Technological Portfolio: How Fairtown Hospital Preserved Results While Degrading Its Older Surgical Robot. *Administrative Science Quarterly*, 000183922311744.

Beane, M., & Orlikowski, W. J. (2015). What Difference Does a Robot Make? The Material Enactment of Distributed Coordination. *Organization Science*, 26(6), 1553–1573.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013.

Bijlsma-Frankema, K., Sitkin, S. B., & Weibel, A. (2015). Distrust in the balance: The emergence and development of intergroup distrust in a court of law. *Organization Science*, 26(4), 1018–1039.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.

Brattström, A., Faems, D., & Mähring, M. (2019). From trust convergence to trust divergence: Trust development in conflictual interorganizational relationships. *Organization Studies*, 40(11), 1685–1711.

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, 82(5), 977–1008.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68(3), 608–624.

Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition*, 7(3), 262–274.

Brynjolfsson, E., & McAfee, A. (2017), July 21. The business of artificial intelligence: What it can—and cannot—do for your organization. *Harvard Business Review*. https://hbr.org/cover-story/2017/07/the-business-of-artifical-intelligence.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530–1534.

Calero Valdez, A., Ziefle, M., Verbert, K., Felfernig, A., & Holzinger, A. (2016). Recommender systems for health informatics: State-of-the-art and future perspectives. In A. Holzinger (Ed.), *Machine learning for health informatics*, vol. 9605 (pp. 391–414). Cham: Springer International.

Candelon, F., Charme di Carlo, R., & Mills, S. D. (2021, October 7). AI-at-scale hinges on gaining a "social license." *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/ai-at-scale-hinges-on-gaining-a-social-license/.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3), 416–431.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The retention of manual flying skills in the automated cockpit: Human factors. *The Journal of the Human Factors and Ergonomics Society*, 56(8), 1506–1516.

Castelvecchi, D. (2016). Can we open the black box of AI? Nature, 538(7623), 20–23.

Chawla, C. (2020). Trust in blockchains: Algorithmic and organizational. *Journal of Business Venturing Insights*, 14, e00203.

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2), 205395171771885.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems, 4691–4697. Presented at the International Joint Conferences on Artificial Intelligence (IJCAI).

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889–6892.

Davies, A. (2018). The WIRED guide to self-driving cars. *Wired Magazine.* https://www.wired.com/story/guide-self-driving-cars/.

Davis, M. S. (1971). That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, 1(2), 309–344.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*, 144(1), 114–126.

Dimoka. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly*, 34(2), 373-396.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697–718.

European Society of Radiology. (2019). What the radiologist should know about artificial intelligence – an ESR white paper. *Insights into Imaging*, 10(1), article number: 44.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70.

Faraj, S., Renno, W., & Bhardwaj, A. (2022). AI and uncertainty in organizing. In M. A. Griffin & G. Grote (Eds.), *The Oxford handbook of uncertainty management in work organizations* (1st ed.), C4.S1-C4.S15. Oxford University Press.

Ferrer, X., Nuenen, T. van, Such, J. M., Cote, M., & Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine*, 40(2), 72–80.

Fleming, P. (2019). Robots and organization studies: Why robots might not want to steal your job. *Organization Studies*, 40(1), 23–38.

Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, 97(4), 62–73.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, 38(4), 1167–1230.

Galperin, R. V. 2017. Mass-production of professional services and pseudo-professional identity in tax preparation work. *Academy of Management Discoveries*, 3(2), 208–229.

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, 31(3), 410–435.

Gillespie, N., Lockey, S., & Curtis, C. (2021). Trust in artificial Intelligence: A five country study. Brisbane, Australia: The University of Queensland and KPMG. https://doi.org/10.14264/e34bfa3.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16(1), 15–31.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: *Strategies for qualitative research*. Aldine.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.

GPT-3. (2020, August 10). A robot wrote this entire article. Are you scared yet, human? *The Guardian*. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614.

Gustafsson, S., Gillespie, N., Searle, R., Hope Hailey, V., & Dietz, G. (2021). Preserving organizational trust during disruption. *Organization Studies*, 42(9), 1409–1433.

Hale, K. (2021, September 2). A.I. bias caused 80% of black mortgage applicants to be denied. *Forbes*. https://www.forbes.com/sites/korihale/2021/09/02/ai-bias-caused-80-of-black-mortgage-applicants-to-be-denied/?sh=26694b3a36fe.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, 57(4), 657–674.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., et al. (2016). *The age of analytics: Competing in a data-driven world*. McKinsey Global Institute.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors: *The Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434.

Hultin, M., & Szulkin, R. (1999). Wages and unequal access to organizational power: An empirical test of gender discrimination. *Administrative Science Quarterly*, 44(3), 453–72.

Iansiti, M., & Lakhani, K. R. (2020). Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. *Harvard Business Review Press*.

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology, & Human Values*, 41(1), 17–49.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. 2(021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635.

Karunakaran, A. (2022). In cloud we trust? Co-opting occupational gatekeepers to produce normalized trust in platform-mediated interorganizational relationships. *Organization Science*, 33(3), 1188–1211.

Kay, M., Matuszek, C., & Munson, S. A. (2015). Unequal representation and gender stereotypes in image search results for occupations. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3819–3828.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410.

Kelly, K. (2017). *The inevitable: Understanding the 12 technological forces that will shape our future.* Penguin.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668.

Lee, P. (2016, May 26). Official Microsoft Blog. Learning from Tay's introduction. https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/.

Leonardi, P. M., & Barley, S. R. (2010). What's Under Construction Here? Social Action, Materiality, and Power in Constructivist Studies of Technology and Organizing. *Academy of Management Annals*, 4(1), 1–51.

Leonardi, P. M., Barley, W. C., & Woo, D. (2022). Why should I trust your model? How to successfully enroll digital models for innovation. *Innovation*, 24(1), 47–64.

Levin, S. T. (2016, September 8). A beauty contest was judged by AI and the robots didn't like dark skin. *The Guardian.* https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people.

Levina, N. & Vaast, E. (2005). The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quarterly*, 29(2), 335-363.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, 23(3), 438–458.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, 32(6), 991–1022.

Lockey, S., Gillespie, N., Holm, D., & Asadi Someh, I. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. *Proceedings of the 54th Hawaii International Conference on System Science.*

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Lou, B., & Wu, L. (2021). AI on drugs: Can artificial intelligence accelerate drug development? Evidence from a large-scale examination of bio-pharma firms. *MIS Quarterly*, 45(3), 1451–1482.

Luccioni, A. S., Akiki, C., Mitchell, M., & Jernite, Y. (2023). Stable bias: Analyzing societal representations in diffusion models. https://doi.org/10.48550/ARXIV.2303.11408.

Lumineau, F., Schilke, O., & Wang, W. (2022). Organizational trust in the age of the fourth industrial revolution: Shifts in the form, production, and targets of trust. *Journal of Management Inquiry*, forthcoming.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40(4), 291–303.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., et al. (2023). *The AI Index 2023 Annual Report*. Stanford University.

McCorduck, P. (2004). Machines who think: *Personal inquiry into the history and prospects of artificial intelligence* (2nd ed.). A. K. Peters.

Metz, C. (2016), March 14. How Google's AI viewed the move no human could understand. *Wired Magazine*. https://www.wired.com/2016/03/googles-ai-viewed-move-no-human-understand/.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, 16(4), 317–342.

Myers, M. D. (1995). Dialectical hermeneutics: a theoretical framework for the implementation of information systems. *Information systems journal*, 5(1), 51-70.

Möhlmann, M., & Zalmanson, L. (2017). Hand on the wheel: Navigating algorithmic managment and Uber drivers', In autonomy, Presented at the Thirty-Eighth International Conference on Information Systems, South Korea (10-13)

Nelson, A. J., & Irwin, J. (2014). "Defining What We Do—All Over Again": Occupational Identity, Technological Change, and the Librarian/Internet-Search Relationship. *Academy of Management Journal*, 57(3), 892–928.

OECD. (n.d.). Recommendation of the Council on Artificial Intelligence. OECD. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st ed.). Crown.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing. *Organization Science*, 32(1), 18–41.

Pawlowski, S. D. & Robey, D. (2004). Bridging user organizations: Knowledge brokering and the work of information technology professionals. *MIS Quarterly*, 28(4), 645-672.

Petit, P. (2007). The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics*, 14(3), 371–391.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Towards human AI hybrids. *MIS Quarterly*, 43(1), iii–ix.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1), 192-210.

Recker, J., Von Briel, F., Yoo, Y., Nagaraj, V., & McManus, M. (2023). Orchestrating Human-Machine Designer Ensembles during Product Innovation. *California Management Review*, 65(3), 27–47.

Ridgeway, C. (2006). *Gender as an organizing force in social relations: Implications for the future of inequality. The declining significance of gender*. Russell Sage Foundation.

Rivera, L. A., & Tilcsik, A. (2016). Class advantage, commitment penalty: The gendered effect of social class signals in an elite labor market. *American Sociological Review*, 81(6), 1097–1131.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.

Sachs, S. E. (2020). The algorithm at work? Explanation and repair in the enactment of similarity in art data. *Information, Communication & Society*, 23(11), 1689–1705.

Saunders, M. N., Dietz, G., & Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? *Human Relations*, 67(6), 639–665.

Sawadogo, P., & Darmont, J. (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, 56(1), 97–120.

Shapiro, A. (2017). Reform predictive policing. *Nature*, 541(7638), 458–460.

Sørensen, O. H., Hasle, P., & Pejtersen, J. H. (2011). Trust relations in management of change. *Scandinavian Journal of Management*, 27(4), 405–417.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, 62(2), 207–218.

Stewart, K. J. (2003). Trust transfer on the world wide web. *Organization Science*, 14(1), 5–17.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., et al. (2016). Artificial intelligence and life in 2030: One hundered year study on artifical intelligence. Report of the 2015 study panel. Staford University. https://creativecommons.org/licenses/by-nd/4.0/.

Teodorescu, M., Morse, L., Awwad, Y., & Kane, G. (2021). Failures of fairness in automation require a deeper understanding of human-ML augmentation. *MIS Quarterly*, 45(3), 1483–1500.

Vaast, E., & Pinsonneault, A. (2021). When digital technologies enable and threaten occupational identity: The delicate balancing act of data scientists. *MIS Quarterly*, 45(3), 1087–1112.

Van de Ven, A. H. (2007). E*ngaged scholarship: A guide for organizational and social research*. Oxford University Press.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, 45(3), 1557–1580.

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4), 404–409.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science*, 33(1), 59–82.

Zetka, J. R. (2001). Occupational divisions of labor and their technology politics: The case of surgical scopes and gastrointestinal medicine. *Social Forces*, 79(4), 1495–1520.

Zuboff, S. (1985). Automate/informate: The two faces of intelligent technology. *Organizational Dynamics*, 14(2), 5–18.

Papers

Paper 1.

Getting AI implementation right: Insights on challenges and solutions from a global survey

Authors:
Rebecka C. Ångström, Michael Björn, Linus Dahlander,
Magnus Mähring, Martin W. Wallin

Status:
Accepted for *California Management Review*

Summary

While the promise of artificial intelligence (AI) is pervasive, many companies struggle with AI implementation challenges. This article presents results from a survey of 2,525 decision-makers with AI experience in China, Germany, India, the United Kingdom, and the United States—as well as interviews with 16 AI implementation experts—in order to understand the challenges companies face when implementing AI. The study covers technological, organizational, and cultural factors and identifies key challenges and solutions for AI implementation. This article develops a diagnostic framework to help executives navigate AI challenges as companies gain momentum, manage organization-wide complexities, and curate a network of partners, algorithms, and data sources to create value through AI.

Keywords: artificial intelligence, innovation, innovation management, innovation focused strategy, change management.

Introduction

Artificial intelligence (AI) transforms how companies compete, interact, and create value with suppliers, employees, and customers (Iansiti & Lakhani, 2020; Brynjolfsson & McAfee, 2014; Haenlein & Kaplan, 2019). But the promises of AI often stand in stark contrast with the many failing AI initiatives that companies experience when embarking on the quest to become data-driven and AI savvy. We know relatively well why companies embrace AI but less about AI implementation efforts beyond oft-repeated examples from highly successful technology firms. To truly realize the potential of AI, we need to build a more solid understanding of how "ordinary firms" – the backbone of most economies – conduct and experience AI implementation.

Understanding such implementation challenges is essential, considering how global spending on AI initiatives reached a whopping $118 billion in 2022 (IDC, 2022) and continues to accelerate while often providing meager results (Tse et al., 2020). Even renowned tech companies struggle to get AI right, as evidenced by IBM's scaling down of its famed Watson technology and Amazon shelving its AI recruitment tool (Jeans, 2020). Behind these famous and infamous examples, recent studies show that many AI implementations are unsuccessful, with 70 percent of companies reporting a minimal impact from AI (Ransbotham et al., 2019) and only 13 percent of data science projects making it into production (VentureBeat, 2019).

The starting point for this paper is therefore simple: While the high-level promises of AI are wide-ranging and partly revolutionary for how companies operate and serve their constituents—for example, through faster and more adaptive communication and knowledge generation, vastly improved analyses and predictions, and streamlined and automated processes previously requiring human judgment (Agrawal et al., 2018; Faraj et al., 2018) —we need to learn much more about the "shop-floor" implementation to deliver on these promises. In particular, we need to understand concrete challenges and solutions that firms encounter and employ when first embarking on AI initiatives and whether and how these differ from when AI becomes more widely adopted and brought to life in organizations. This means that we need to look beyond the success stories of the likes of Facebook, Google, Tencent, and Microsoft, which may blind us to many of the challenges most companies face.

To tackle this issue, we surveyed 2,525 decision-makers from organizations currently implementing AI in five countries: China, Germany, India, the UK, and the US. We distilled insights from more experienced ("AI Experienced") firms and less experienced ("AI Newcomers") in a wide range of industries in these global geographies. To add further insight, we conducted 16 interviews with AI implementation experts (executives with extensive AI implementation experience) in Sweden and the UK. Our findings are highly relevant for any manager shouldering the task of diffusing effective AI practices into the wider organization.

What we do and do not know about AI implementation

Propelled by advances in computational power, programming science, and access to large data sets (Faraj et al., 2018; Nilsson, 1998; Waardenburg et al., 2022), AI technologies are reaching a crucial stage of development (Faraj et al., 2018) in areas such as machine learning, pattern recognition, computer vision, and natural language processing (Zhang et al., 2022) to name but a few. AI is commonly defined as "the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity." (Rai et al., 2019, p. iii). Like other information technologies, AI will transform work by taking over tasks previously carried out by humans, allowing humans to focus on more complicated and rewarding tasks. Unlike most information technologies, however, AI can augment human judgment and blend with human activities in entirely new ways and diverse settings (e.g., human-AI-surgery, semi-autonomous drones, etc.) (Faraj et al., 2018; Rai et al., 2019). AI solutions are already changing work, transforming expertise, reshaping occupational boundaries, and introducing new methods of control and decision-making. These developments fuel predictions that investments in AI will continue to increase and that productivity increases will subsequently follow (Brynjolfsson et al., 2019).

According to scholars and thought leaders, however, firms need to rethink how they organize and operate in order to reap these benefits from AI. For example, firms must change their operating model from one in which they deliver a specific product, service, or solution to one in which they design a "software-automated, algorithm-driven digital 'organization'" (Iansiti & Lakhani, 2020). Work practices must be re-engineered in this transformation as workflows are broken down into smaller tasks corresponding to individual AI algorithms' relatively limited capabilities (Agrawal et al., 2018; Kolbjørnsrud et al., 2017; Ross, 2018). Firms will also need to put in place new roles (AI-business translator, data scientist) and invest in setting up new units, such as "AI factories" with data pipelines, algorithm development, experimentation platforms, and related software architectures (Iansiti & Lakhani, 2020; Vaast & Pinsonneault, 2021).

In this new landscape, where AI as a rapidly evolving general-purpose technology will continue to find new application areas, managers are likely to become less decision-makers and more curators of portfolios of algorithms and data flows, and employees that possess both AI and domain-specific skills will be particularly sought-after (Iansiti & Lakhani, 2020; Lou & Wu, 2021; Shrestha et al., 2019). In parallel, concerns regarding the liability, trustworthiness, and ethical usage of AI algorithms, including risks for privacy violations through AI and data, are being raised (Berente et al., 2021; Crawford, 2021; European Union Agency for Fundamental Rights., 2020; Iansiti & Lakhani, 2020; Rai et al., 2019), alongside proposals for regulations to address how and where AI should be implemented (Burt, 2021).

The many distinct characteristics of AI technologies make it essential to consider the AI implementation challenge as different from implementing established information technologies, such as ERP systems, CRM systems, HR systems, and productivity and communication software. These are typically stable and well-integrated software products with highly controlled (and relatively infrequent) release schedules and with high reliability and predictability in their functionality. This contrasts the often granular, dynamic, tentative, and incomplete functioning of combinations of AI algorithms, which exhibit much greater volatility in evolution and use. Indeed, the potential, use, consequences, and expertise needed for AI are quite different from traditional IT, and it thus becomes essential to understand how AI can be implemented. That is, we must figure out how to design, configure, and deploy AI technologies and adapt organizational structures and routines to realize the technology's potential while accommodating different demands (Aanestad & Jensen, 2016; Asatiani et al., 2021; Avgerou & Bonina, 2020). However, literature on AI implementation in organizations is scarce. A few studies have highlighted the importance of AI-relevant competencies (e.g., data scientists with deployment-oriented skills or domain experts that can make productive use of data) (Davenport & Malone, 2021; Ross, 2018) and stressed risks arising from the under-performance of AI technologies (Strohm et al., 2020) and the lacking availability and quality of data (Cabitza et al., 2020). While these and other studies highlight the usefulness of attending to technological, as well as organizational and cultural (people) challenges (Benbya et al., 2020, 2021), much remains to explore to provide actionable knowledge for managers charged with leading AI implementation initiatives. We thus set out to survey a large sample of firms to discover specific challenges and solutions associated with AI implementation.

Data and research design

We combined expert interviews with a survey of white-collar decision-makers. As a first step, we conducted 16 in-depth interviews with AI implementation experts in Sweden and the UK in sectors such as IT, telecommunications, banking and finance, insurance, fintech, and the public sector. The experts represented different roles, such as Chief Technology Officers (CTOs), Chief Digital Officers (CDOs), government experts, and suppliers of AI and platforms, all with extensive experience in implementing AI. The interviews revealed that implementing AI is a complex and laborious process. They confirmed the relevance of capturing challenges not only around the technology but also including a broad spectrum of organizational and cultural issues. Based on these insights, we developed an online survey to probe into the challenges of AI implementation[5] in the three distinct domains: technological, organizational, and cultural.

Next, we collected survey data from 2,525 decision-makers in China, Germany, India, the UK, and the US with experience in implementing AI. We sampled respondents from panels of online business professionals adhering to ESOMAR[6] quality controls. In addition, we assessed the number of surveys taken, response patterns, number of screen-outs, and real-time digital fingerprinting against fraudulent behaviors. We divided surveys equally among countries to reach a target quota of 500 decision-makers with

[5] In our pre-study work we found that "AI" was sometimes viewed as futuristic and almost unachievable. As such, we expanded the focus of the survey to also include "advanced analytics", a label that resonated well with many experts. Specifically, we prompted survey respondents with the following statement: "We are interested in understanding the adoption and implementation of technologies that draw on data in order to generate new insights or to automate processes, such as AI, machine learning and advanced analytics. This includes sophisticated applications such as prediction models, data pattern recognition, digital assistants, image analysis software, speech and face recognition systems, just to mention a few."

[6] European Society for Opinion and Marketing Research, https://esomar.org

AI implementation experience from each country. On a per-country basis, quotas were further subdivided to survey at least 250 technical managers responsible for introducing AI technologies into their companies and 250 operational managers tasked with using AI in their operational processes. To reach these quotas, we sampled 10,024 full-time professionals in companies with at least 100 employees, of whom 6,781 were white-collar decision-makers. A mix of panel sources was used to avoid country-specific effects and biases. Each country's targeted audience was identical, and panels had no pre-existing focus on AI that could influence the sample.

We asked respondents to share information about their organizations' historical and current initiatives to implement AI. Survey questions addressed the challenges encountered during AI implementation, strategies to overcome those challenges, and plans for investment and hiring. We pre-tested the survey for both understandability and translation of the different languages. On average, it took respondents 18 minutes to complete the survey.

We also conducted additional analyses to distinguish two subcategories of firms, AI Experienced firms and AI Newcomers, to better understand how the challenges of AI implementation vary with experience. We defined AI Experienced firms as having at least one fully implemented AI system and AI Newcomers as actively pursuing AI solutions but not yet having completed their first AI implementation effort.[7] Given the early stage of AI implementation, we opted for an inclusive understanding of being AI Experienced to capture companies ahead of the curve (note that three out of four firms initially approached were deselected due to not being active in AI implementation). This means that our study reaches well beyond the all-known "AI superstars", such as the Big Tech firms, to capture AI challenges and solutions among the many. Below, we report on commonalities and differences in implementation challenges between AI Experienced firms and AI Newcomers.

[7] Specifically we asked "Has your company implemented any kind of AI or Advanced Analytics tools?"

The technological, organizational, and cultural challenges of implementing AI

Our expert interviews taught us various challenges associated with AI implementation, from lacking visibility of available data to employees preferring to trust their intuition over data analytics. Dividing the AI implementation challenges into three domains: technological, organizational, and cultural, we analyzed each challenge based on our survey data. This revealed that nearly all (ninety-nine percent) of our respondents had encountered at least one challenge in one of the domains as they implemented AI. Ninety-one percent had encountered challenges in all three domains. As indicated in Figure 1, AI Experienced firms and AI Newcomers face many similar challenges (see all challenges where only the grand mean is displayed). As we compared organizations based on the maturity of their AI initiatives, an overarching insight emerged: gaining experience does not lessen the trials and tribulations of AI implementation. Instead, increased maturity comes hand-in-hand with new challenges and, in some cases, exacerbates existing ones.

Figure 1. Technological, organizational, and cultural challenges

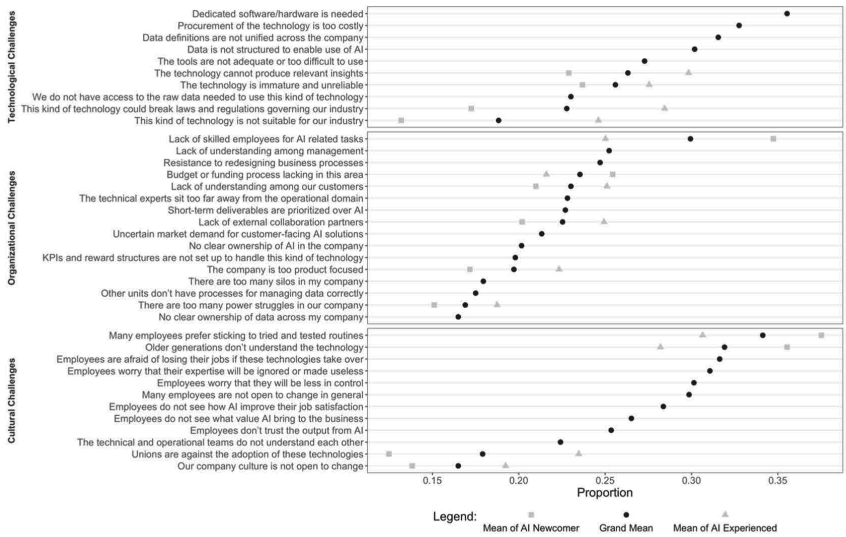


Figure 1 shows the compound results for the full sample (both AI Experienced and Newcomers) along the dimensions of technological, organizational, and cultural challenges, as well as the identified statistically significant differences between the two groups of firms. The first pattern that emerges is that some challenges are more prevalent than others, and many of these are of a technological or cultural rather than organizational nature. The most frequently reported technological challenges are the need for dedicated software and hardware and related investments and data management issues. Indeed, many technological challenges can be traced back to data: data not being available, data not being structured for the desired use, and data definitions not being unified across the company. One of the AI implementation experts we interviewed explained that without a unified language, structured data assets, and proper coordination across the company, it becomes nearly impossible to implement AI (see Box 1).

Box 1: AI leadership builds on a unified view of data.
When Bank A (an AI Experienced firm) started its transformational journey to becoming AI-driven, different units handled data locally. The lack of coordination between units led to several challenges. For instance, each unit had developed its terminology for different information, which led to confusion when communicating across units. Another challenge was the lack of visibility on what data were available within each unit. Searching for data became time-consuming, often achieved by informally asking around within one's network. In turn, this practice increased the chance of misunderstandings, leading to late, manual, and costly dataset corrections. The bank realized that there were significant efficiency gains to be made by becoming more organized and that they would need to make data management a prerequisite if they wanted to advance in AI development. They created a unit with the sole purpose of helping the bank manage data as an asset. The unit is now in charge of the bank's strategic development in terms of data, including developing its data governance model and delegating data ownership within the bank.

Interestingly, we found that AI Experienced firms are more, rather than less, likely to face challenges with bending AI technologies to their will, reporting a lack of fit with industry-specific needs and problems with leveraging the technology to get at salient insights. Experienced firms are also more likely to foresee the risk of breaking laws and regulations. (One of our experts pointed out that laws and regulations hamper innovation and learning, as complying with them is costly and restraining.) We detect a growing appetite for more advanced solutions as you gather experience: Experienced firms push the envelope regarding ambitions and complexity, running into restrictions concerning technology and data that Newcomers have yet to face. They make progress, but life does not get easier. For example, AI Experienced firms are more likely to have resolved issues around data ownership, partly by promoting a data-driven workflow across organizational boundaries.

The most frequently mentioned organizational challenge (by some margin) is the lack of adequately trained employees. Still, the lack of understanding of AI among managers, as well as among customers, also stands out. This suggests that despite the extensive investments already taking place, considerably more investment in skills development can be expected, also in Experienced firms. One of our experts shared how leaders within a data-savvy firm were reluctant to accept data analytics in decision-making. The firm had a data scientist embedded within each product development team, and one of the tasks of the data scientist was to run an A/B test for new ideas. However, when the test showed that the idea would not measure up, the data scientist faced an escalating conflict with the product owner. The product owner even preferred statistically ambiguous results since they afforded "interpretive flexibility" and allowed for acting on gut feeling.

While building skills and understanding takes time, we see that Experienced firms have made some headway: Newcomers experience more challenges concerning employee skills. They also report a higher incidence of challenges related to AI and data governance (lack of clear ownership). The good news here is that effort pays off: Experienced firms are gaining ground in the form of reduced incidence of some organizational challenges. However, this is also a story of perseverance, where challenges evolve but remain as complexity grows.

Finally, companies also struggle with cultural challenges. The most common occurrences are the combination of different fears and concerns that employees carry concerning AI (job loss, loss of expertise, and autonomy), in combination with inertia in routines and a perceived generational gap in technology understanding. Generally, companies report a higher prevalence of cultural challenges than organizational ones. Moreover, Newcomers report significantly more cultural challenges than Experienced firms: employees are more fearful of AI, and there is stronger resistance to the technology, sticky routines represent a more frequent hurdle, and employees fail to see what value AI can bring to the business and to job satisfaction. As one expert explained, working with organizations that lack a data-driven mindset, it is a psychological challenge to get employees to realize that their intuition, often built on many years of experience, may be wrong and that they instead need to trust the data. In contrast, Experienced firms more frequently report that trade unions oppose AI adoption (although overall, this is a lesser challenge).

In sum, cultural and technology challenges are front and center as companies enter the AI arena. Still, Experienced firms do a better job addressing them when new challenges come to the fore.

Solutions to advance AI implementation

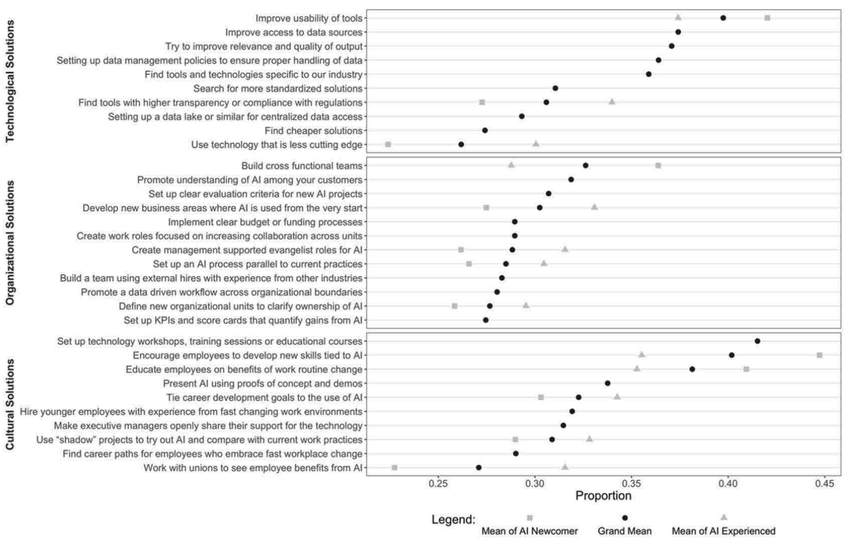
Having established the challenges that AI Experienced firms and Newcomers experience, we now turn to the solutions that managers employ to address the challenges of AI implementation. Here also, we find many solutions to be shared for Experienced firms and Newcomers. Among technological solutions, improving the usability of AI tools was most often mentioned, followed by improved access to data, enhancing the quality of output from algorithms, improved data management, and finding suitable tools. Note that all these solutions focus on the core work of getting algorithms to operate in alignment with expectations to produce value-adding results. Most solutions we identified were employed in roughly equal measures. Still, we detected significant differences concerning tool selection, with Newcomers more inclined to prioritize standardized solutions and Experienced firms more often looking for tools that provide transparency and regulatory compliance. Again, this suggests that as firms become more experienced, they are hitting new hurdles, such as algorithmic opacity (driving the need for transparency) and more advanced solutions pushing privacy and legal boundaries (driving the need to ensure regulatory compliance).

As with challenges, firms seemed to pay less attention to organizational solutions than technological and cultural solutions. Also, the spread in popularity across different organizational solutions is minor (see Appendix 2 for details). In other words, a broad range of solutions is similarly crucial. Many of these solutions focus on organizing and governing AI innovation activities, including creating new roles, building cross-functional teams, and developing control measures such as budgets and evaluation criteria. Here, we also identify two areas where Newcomers and Experienced firms differ: Newcomers focus more on building cross-functional teams.

In contrast, Experienced firms focus more on promoting a data-driven workflow across organizational boundaries and creating AI evangelist roles. Both these latter solutions are typical for firms that have already made headway and gained momentum. Before creating an organization-wide AI evangelist role, you need early wins to build on and people with experience to share; otherwise, the evangelist can come across as a false prophet. Similarly, data-driven workflows across organizational boundaries require involving customers and partners and need to build on a foundation of early successes and honed capabilities.

Finally, the cultural solutions are focused on skills development and change management. Three stand out: workshops and training, promoting employee skills development on AI, and changing work routines. But also proof of concept studies and demos, and a slew of other activities are used to push the advancement of AI. Further stressing the importance of training is that Newcomers focus even more on AI skills development (see Figure 2).

Figure 2. Technological, organizational, and cultural solutions



Key insights for managers

Experience Breeds Ambition, Complexity, and Continued Implementation Challenges

It would be natural to assume that AI Experienced firms would enjoy a smoother ride with fewer and less severe challenges than AI Newcomers. However, our data suggest that challenges persist and even grow in complexity. Consider the case of a supplier in the automotive industry. Initially, their major struggles when implementing an autonomous vehicle AI system were developing effective algorithms and attracting skilled people. As their AI implementation matured, they discovered the need to integrate databases, which led to challenges in coordinating people in new ways across departments. So, challenges did not disappear, but their nature evolved, and complexity grew as distinct technological challenges became organizationally entangled.

Managing this increasing complexity includes several lines of action, such as experimentation (proof of concept initiatives) and learning, relentless focus on skills development, and gradually adding organizational and governance solutions (including roles and units with ownership over certain AI tasks and processes). The usefulness of pilots is emphasized in our survey and by AI implementation experts, who see them as a starting point for implementing AI and building an understanding of AI capabilities within the organization. Such pilots are often "low-hanging-fruit" with well-defined application areas that rely on available data from one or just a few sources. One of our experts shared such an experience with developing an algorithm for predictive maintenance of power tools. A product team in charge of developing the power tool decided to use AI to predict when the tool needed maintenance. The indicator for maintenance was when the tool started to make noises indicating wear and tear. The team used microphones to record the sound of the power tool and taught the algorithm the difference between a well-functioning tool and one needing maintenance. The development was straightforward and did not involve any other team.

However, as AI Experienced firms expand to more advanced application areas, this strains the technology and the data at hand. For example, as Iansiti and Lakhani observe, as the sophistication of AI use increases, the need for policies and architecture (Iansiti & Lakhani, 2020) and centralized data management grows. For example, the need to access data from outside sources is underscored by our finding that AI Experienced firms, to a larger extent, promote a data-driven workflow across organizational boundaries. Venturing beyond well-defined use cases and including customers and partners, however, drives complexity in solutions and pushes technological as well as legal and regulatory boundaries.

So how do Experienced firms address the technological complexity risks associated with being ahead of the curve? Perhaps surprisingly, we find some support for firms graduating to more advanced AI solutions often looks for simpler tools that are less groundbreaking—that is, they focus on finding the point where they can be "leading edge" in AI applications without being "bleeding edge" in AI tool selection, finding the right tool for each job.

In sum, experience drives ambition, which drives complexity. The trick is to constantly balance complexity against capabilities and grow your people skills as rapidly as possible.

Invest in People. Then Invest Some More

A strong pattern across challenges and solutions, and across Experienced firms and Newcomers, is the need to attract and develop people. While rapid technological advances drive AI adoption, AI implementation is not a technology problem best solved by a few (or many) dedicated data scientists. It is an organizational transformation and value-creation challenge driven by technology and data solutions, people, and supporting organizational arrangements in concert. Our respondents report many cultural and organizational challenges equally important for AI to gain a foothold and garner momentum (see Figure 1). Recall that ninety-one of our informants reported challenges across all surveyed categories, technology, organization, and culture. People-related issues stand out both amongst cultural and organizational challenges. So, investing in AI means investing in people.

For example, consider the major European bank that used AI to become "data-driven": the initiative took off only when they connected the technology with people and purpose (see Box 2). Similarly, the head of the analytics team at another European bank shared the insight that when people don't understand the substance of AI—what AI really entails—they gravitate toward quick fixes with unrealistic expectations of results. You can't just hire a couple of data scientists and expect wonders. You need people who know what data can be made available to feed the AI algorithms and how to interpret and make use of the results. On the other hand, brilliant data scientists who do not understand the operational context will have a hard time creating true value. A bank representative told us that data scientists with dazzling tech skills, cutting-edge statistical acumen, and compelling academic CVs are often hampered by a lack of understanding of the business and how to create value. Hence the need for broad skills development and cross-functional teams, as shown in our survey results.

Box 2: Technological awareness is not enough.
Bank B (an AI Experienced firm) started its journey toward becoming data-driven with a top-down approach. Part of the bank's strategy was to set up an information governance model, where appointed staff became data owners responsible for the data in the bank. Data ownership could include being responsible for storage, access, and quality assurance. The approach was partly successful at creating awareness, but it did not yield any results regarding activities or initiatives. The bank then adopted a more agile working method, focusing on smaller pilot projects. This was done by identifying and approaching units with indicators of problems related to data management, for instance, being fined by or receiving reminders from the financial regulatory authority. Together with these units, the bank created pilot projects that could also work elsewhere in the organization. This time, the employees acted quickly and didn't question the purpose of the information governance model. The bank recognized these changes as an effect of the increased awareness created by the workshops in the first part of the initiative.

A striking difference between AI Experienced firms and AI Newcomers is that the former has made considerably more headway in their people skills and attitudes. AI Experienced firms less often report a shortage of AI-skilled employees, fewer challenges handling employees' fear of AI, and a smaller generational gap regarding employee preparedness. They are also more proactive in working with unions to frame AI as an opportunity for employees rather than a threat to their livelihood. Where Newcomers struggle to find the right expertise and motivate employees, Experienced firms have a more refined management understanding of AI.

Overall, the focus on learning is striking: proof of concepts and demos, education/training/workshops, and encouraging personal growth all rank highly. This suggests that for sustained AI adoption and AI-driven change, companies must create an informed, interested, and engaged workforce able and willing to work routinely with AI-based process improvements and solutions development. This means that investments in AI need to be combined with dedicated investments in people who can grow and remain innovative as AI technologies evolve. Many experts highlighted the need to stimulate employees to learn about data and AI. For instance, one of our experts shared that to encourage openness to learning and countering resistance, their technology firm encourages continuous learning and promotes a growth mindset. Another of our experts argued that the point of training is not to turn everyone in the organization into data scientists but that everyone needs to understand the basic concepts of data and AI, as well as develop an understanding of data and AI that cannot be taught through a traditional digital course but must be rooted in personal experience.

Shift your mindset: From software to algorithms and from stable to dynamic governance

Creating an AI-driven organization places new challenges on how to manage digital technologies. Even companies with excellent IT expertise need to adapt to the world of dynamic algorithms voraciously hungry for data. The initial implementations of AI applications to address specific challenges is a good start, but it is a far cry from creating an AI-driven organization. It is also very different from managing standard IT resources, as it needs constant and iterative attention from both technical and operational domain experts after deployment (Lou & Wu, 2021). Regular IT systems—think of customer relationship management (CRM) systems or enterprise resource planning (ERP) systems—are expected to perform reliably over long periods, supported only by planned, infrequent updates. Such standard software can be developed by vendor organizations and deployed in a similar fashion across many organizations.

In contrast, to ensure that each AI application delivers value, not only does development need to be done in close coordination with operations, but the algorithms also need constant care to ensure that it performs as intended. This highly specific oversight is needed partly because AI applications are often trained on data streams from the same settings where they will be deployed. Identifying what data to include and how those data should be interpreted calls for advanced, domain-specific knowledge of business goals, existing processes, and the context in which data are derived and used. This can also include data from other AI applications creating an intricate ecosystem of data, dynamic, reusable, and modifiable algorithms, and resulting solution bundles. As AI applications evolve, they can improve their performance but also deviate or fail, requiring adjustment or retraining. This, again, calls for cross-disciplinary expertise from both the technical and operational domains.

Having only a few AI applications, this is not necessarily a problem, but as the number of AI applications grows, so does the complexity of the algorithmic ecosystem in which they are implemented. Developing and curating algorithms will call for much hands-on work—a process that can be both time and resource-consuming and sometimes unsustainable (see Box 3). Furthermore, having multiple, interdependent AI applications—each contributing to different processes and drawing from different datasets—puts high demands on companies to organize data in a manner that avoids conflicting decisions and processes, both as datasets are modified and added and as AI applications evolve, collaborate, and depend on one another.

Having more AI applications in place not only increases the complexity of implementation and management of the technology but also impacts the operational domain competence. One of our experts compares the difference between AI and traditional IT by explaining that AI comes much closer to the employee and, in practice, becomes an extension of that person's competence. Instead of supporting an employee's activity, the AI performs part of it. And the employee needs to understand what the AI has done and why and incorporate this knowledge into their domain expertise.

Box 3: Towards a single platform: Standardization can mitigate escalating maintenance costs.
At Technology Firm A (AI Experienced firm), a team of data scientists supported a wider organization consisting of numerous accounts to develop AI applications models for operational work. At first, the number of requests for the team's service was modest, and little coordination was needed. However, as the number of requests grew, the team soon realized that several data scientists were working on applications for different accounts that could be merged into one. Beyond duplication of work, the greater number of applications resulted in higher maintenance costs. To address these problems, the data scientist team decided to standardize the AI applications. They appointed two team members as gatekeepers responsible for investigating all novel requests and synchronizing work where needed.

Climbing the AI implementation ladder also becomes markedly costlier when applications involve and rely on customers. As the appetite to create more value grows, companies need to consider the wider ecosystem and exploit opportunities beyond the company's borders. Companies cannot bet on finding the necessary data only on the inside of their organization, and securing commitment outside the home organization adds complex new people challenges.

Diagnosing AI implementation activities on different levels

The above insights point towards the need to manage emerging complexities as AI implementation efforts, which often start small, become more pervasive and sprawling, ultimately reaching outside the organization. Figure 3 provides three sets of questions managers must ask as the company gains AI maturity. The first set of questions will help managers gain momentum with AI initiatives in their local setting. The second set prepares managers for the internal complexities arising when the organization's volume, ambition level, and complexity of AI implementation initiatives increase. The third set identifies critical issues managers will likely face as curating a growing AI ecosystem of partners, algorithms, and data sources become increasingly critical.

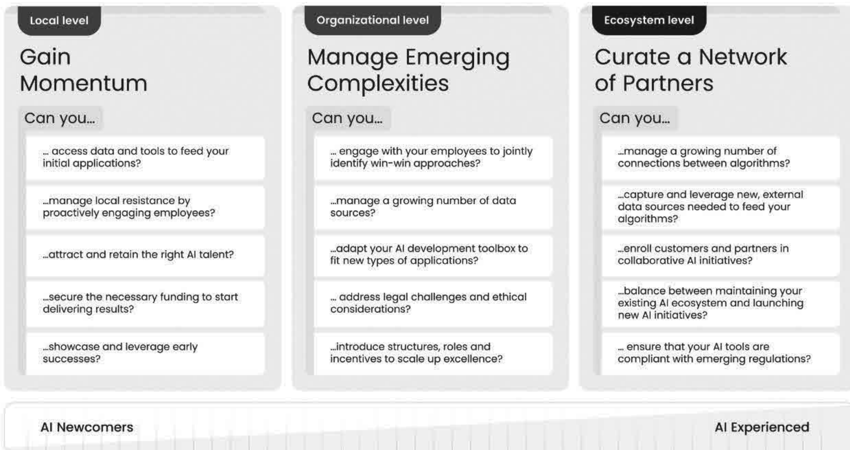
Gaining local momentum

When companies implement AI, a localized and contained approach is often appropriate to showcase that AI can solve a relevant business problem and build experience and internal expertise. Companies will rely on key individuals driving AI, despite an ambition to simplify and de-risk. Moreover, the range of people challenges is broad: from securing tech expertise to building business expertise in AI development, overcoming resistance, building trust in AI across business areas and roles, and not least, building a broader management understanding of what AI really is. Our findings are clear: local pilots, experiments, and other activities that speed up the cycle of action-evaluation-learning help AI implementation gain momentum. But overcoming people challenges alone is not enough. Managers must also secure access to quality data to feed initial applications and ensure that investments in adequate tools are made from the start—before taking on more complex tasks.

Managing organizational complexities

As firms become savvier, more advanced algorithms and solutions require even better skills and expose shortcomings in available technologies; complexity increases, and technological challenges continue. Experienced firms wield a broader and more creative set of tactics to overcome the challenges and advance their AI practice, sometimes counterintuitively seeking simpler tech to avoid getting stuck and relying on "simple rules" to manage complexity. Moving towards organization-wide AI implementation, managers need to organize work to support AI adoption, such as centralizing data access and promoting data-driven workflows. To handle data, it is vital to set clear processes, goals, and ownership for data management and to secure data quality and accessibility across organizational units, as well as ensure regulatory compliance. As the discussion on trustworthy and ethical AI will likely continue to place demands on technology use, forward-looking firms must consider this early when forming a data management strategy.

Figure 3. A diagnostic test for AI implementation



Curating a growing ecosystem

Becoming a truly AI-driven organization requires curating and nurturing a sprawling and complex web of algorithms, data, and partners to ensure that AI solutions are effective upon deployment and are continuously fine-tuned to their missions. As AI Experienced firms approach AI leadership status, they must push through complexity while balancing their growing ambitions and installed base. Building AI functionality that engages customers and partners requires advanced skills in relationship management. Setting up so-called "AI factories" (Iansiti & Lakhani, 2020) with supplementary data pipelines, experimental platforms, and software architectures is painstaking work for any company, often executed in parallel with delivering on previous commitments to stakeholders and customers. As the complexity of the AI application ecosystem grows, new challenges appear. This includes realizing the shortcomings of existing technology and finding tools that support the organization's evolving needs. It also includes responding to the increasing demand for trustworthy and ethical AI by finding tools that are transparent and compliant with emerging regulations, as well as collaborating with an increasing multitude of stakeholders such as customers, unions, and industry organizations.

Conclusion: A journey without end

If you are an AI Newcomer, a split vision is required. Initially, it is imperative to focus on concrete and delimited use cases with clear value propositions and limited complexity in algorithm development, data access, organizational scope, and risk management. At the same time, you must also prepare to manage emerging complexities by proactively investing in data management capabilities, a more fine-tuned portfolio of AI tools, broad people involvement and skills development, deep AI expertise, and nuanced management understanding.

If you are already AI Experienced, your next-level challenge will likely involve dealing with an increasingly complex ecosystem of algorithms, data, solutions, and partners. Add changing work processes and negotiating boundaries and responsibilities with employees to this mix. Some AI Experienced firms will stumble as they navigate the increasing complexity inherent in mastering the integration of AI within the firm and across partner organizations. For more experienced firms, a key insight for navigating the challenges of AI implementation is that you will never be fully AI-proficient. Instead, new and different technological, organizational, and cultural challenges will conspire to play tricks on even the most successful organizations. From its often-disorienting beginnings to the unexpected challenges of growth and maturity, AI implementation is likely to be a journey without end.

References

Aanestad, M., & Jensen, T. B. (2016). Collective Mindfulness in Post-Implementation is Adaptation Processes. *Information and Organization*, 26(1–2), 13–27.

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. *Harvard Business Review Press.*

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical Envelopment of Artificial Intelligence: An Approach to Organizational Deployment of Inscrutable Artificial Intelligence Systems. *Journal of the Association for Information Systems*, 22(2), 325–352.

Avgerou, C., & Bonina, C. (2020). Ideologies Implicated in IT Innovation in Government: A Critical Discourse Analysis of Mexico's International Trade Administration. *Information Systems Journal*, 30(1), 70–95.

Benbya, H., Davenport, T. H., & Pachidi, S. (2020). Artificial Intelligence in Organizations: Current State and Future Opportunities. *SSRN Electronic Journal.*

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications For Information Systems Research. *Journal of the Association For Information Systems*, 22(2), 281–303.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, 45(3), 1433–1450.

Brynjolfsson, E., & Mcafee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies (First Edition). W. W. Norton & Company.

Brynjolfsson, E., Rock, D., & Syverson, C. (2019). Artificial Intelligence and the Modern Productivity Paradox: A Clash Of Expectations And Statistics. In the Economics of Artificial Intelligence: An Agenda (Pp. 23–57). *The University of Chicago Press.*

Burt, A. (2021, April 30). New AI Regulations are Coming. Is Your Organization Ready? *Harvard Business Review.*

Cabitza, F., Campagner, A., & Balsano, C. (2020). Bridging the "Last Mile" Gap Between AI Implementation and Operation: "Data Awareness" That Matters. *Annals Of Translational Medicine*, 8(7), 501–501. H

Crawford, K. (2021). Atlas Of Ai: Power, Politics, and the Planetary Costs of Artificial Intelligence. *Yale University Press.*

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. *Harvard Data Science Review.*

European Union Agency for Fundamental Rights. (2020). Getting the Future Right : Artificial Intelligence and Fundamental Rights: Report. Publications Office. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and Organizing in the Age of the Learning Algorithm. *Information and Organization*, 28(1), 62–70.

Haenlein, M., & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4), 5–14.

Iansiti, M., & Lakhani, K. R. (2020). Competing in the Age of AI: Strategy and Leadership when Algorithms and Networks Run the World. *Harvard Business Review Press.*

Jeans, D. (2020). Companies Will Spend $50 Billion on Artificial Intelligence This Year with Little to Show for it. *Forbes Magazine.* https://www.forbes.com/sites/davidjeans/2020/10/20/bcg-mit-report-shows-companies-will-spend-50-billion-on-artificial-intelligence-with-few-results/

Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2017). Partnering With AI: How Organizations can win Over Skeptical Managers. *Strategy & Leadership*, 45(1), 37–43.

Lou, B., & Wu, L. (2021). AI on Drugs: Can Artificial Intelligence Accelerate Drug Development? Evidence From a Large-Scale Examination of Bio-Pharma Firms. *MIS Quarterly*, 45(3),

Nilsson, N. J. (1998). Artificial Intelligence: A New Synthesis. Morgan Kaufmann.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms: Towards Human AI Hybrids. *MIS Quarterly*, 43(1), Iii–Ix.

Ransbotham, S., Khodabandeh, S., Fehling, R., & Lafountain, B. (2019). Winning With AI. *MIT Sloan Management Review* and Boston Consulting Group. Https://Sloanreview.MIT.edu/Projects/Winning-With-AI/

Ross, J. (2018). The Fundamental Flaw in AI Implementation. *MIT Sloan Management Review*, 59(2), 10–11.

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational Decision-Making Structures in the Age of Artificial Intelligence. *California Management Review*, 61(4), 66–83.

Strohm, L., Hehakaya, C., Ranschaert, E. R., Boon, W. P. C., & Moors, E. H. M. (2020). Implementation Of Artificial Intelligence (Ai) Applications in Radiology: Hindering and Facilitating Factors. *European Radiology*, 30(10), 5525–5532.

Tse, T., Esposito, M., Takaaki, M., & Goh, D. (2020). The Dumb Reason Your AI Project Will Fail. Harvard Business Review Digital Articles, 2-5. *Harvard Business Review Digital Articles*, 2020(2–5).

Vaast, E., & Pinsonneault, A. (2021). When Digital Technologies Enable and Threaten Occupational Identity: The Delicate Balancing Act of Data Scientists. *MIS Quarterly*, 45(3), 1087–1112.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In The Land of the Blind, the One-Eyed Man is King: Knowledge Brokerage in The Age Of Learning Algorithms. *Organization Science*, 33(1), 59–82.

VentureBeat (2019). Why Do 87% Of Data Science Projects Never Make It into Production?. Https://Venturebeat.Com/Ai/Why-Do-87-Of-Data-Science-Projects-Never-Make-It-Into-Production/

Worldwide Spending on AI-Centric Systems Will Pass $300 Billion By 2026, According to IDC. (2022, September 12). International Data Corporation (IDC). https://www.idc.com/getdoc.jsp?containerId=prUS49670322

Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J., & Perrault, R. (2022). The AI Index 2022 Annual Report. AI Index Steering Committee, Stanford Institute For Human-Centered AI, *Stanford University*. https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf

Online appendices

Appendix 1. Construct validity: Assessing whether AI Experienced and AI Newcomers are distinctly different

Our differentiation between AI Experienced and AI Newcomers is based on a self-assessment of their level of AI maturity (resulting in 1244 AI Experienced and 1281 AI Newcomers). Specifically, we defined AI Experienced as having at least one fully implemented AI system and AI Newcomers as having only partly implemented or were currently in the process of implementing an AI system. To ensure construct validity, we analyzed two alternative and complementary measures of AI maturity and compared them to our main measure of AI maturity. This hinges on AI Experienced using (1) more advanced AI work practices and (2) more advanced data sharing and use practices. These results are summarized below.

AI Experienced use more advanced AI practices than AI Newcomers

We captured how respondents assessed their AI-related work practices, using five items (below) that experts consider state-of-the-art for AI: (Iansiti & Lakhani, 2020; Panetta, 2019)

- We use proof of concept or pilot projects extensively to test AI
- AI projects are put into production and used as best practice on a regular basis
- AI projects always have an executive sponsor and a dedicated budget
- AI is considered for use in all new projects, products and services
- We have a clear strategy for AI related data management

Each item was scored on a seven-point Likert scale, ranging from "strongly disagree" to "strongly agree." We constructed the degree of AI advanced practices variable by averaging the scores assigned to these statements. The Cronbach's alpha for this measure was 0.81, well above the accepted threshold of 0.7. We then compared our measure of AI Experienced and AI Newcomers. We found that AI Experienced scored significantly higher on this measure (5.82 compared with AI Newcomers 5.31, t-test significant at 1% level). This shows that AI Experienced, to a much greater extent, have a higher level of maturity when it comes to working with AI than AI Newcomers.

AI Experienced adopt more advanced data sharing and use practices than AI Newcomers

- We also captured how respondents assessed their corporate data sharing practices using four items.
- My company uses data published by other companies in order to improve our own AI activities.
- My company shares data so that other companies can use it in their business activities.
- My company freely shares some data with no restrictions or associated costs.
- My company has a good understanding of licenses governing the use of data.

Like above, we constructed the degree of AI data sharing variable by averaging the scores assigned to these statements. The Cronbach alpha was .75 for the measure. AI Experienced scored significantly higher on this measure than AI Newcomers (5.71 compared with 4.95, significant at the 1% level), underscoring that AI Experienced are considerably more advanced in their strategies for sharing and using data than AI Newcomers.

References

Iansiti, M., & Lakhani, K. R. (2020). Competing in the Age of AI: Strategy and Leadership when Algorithms and Networks Run the World. *Harvard Business Review Press*.

Panetta, K. (2019, February 5). The CIO's Guide to Artificial Intelligence. Gartner.com. Https://Www.Gartner.Com/Smarterwithgartner/The-Cios-Guide-To-Artificial-Intelligence

Appendix 2. Sample and response rate

We conducted an online survey of 2,525 white-collar decision-makers in China, Germany, India, the UK, and the US. A multi-sourcing online recruitment model was used to achieve diversity, coverage, and consistency while minimizing bias. Members in the panels utilized were recruited online and via referrals and active registration, using double opt-in, and only panels that adhere to the ESOMAR quality controls were included. In addition, algorithmic checks for factors including the number of surveys taken, response patterns, number of screen-outs, and real-time digital fingerprinting against fraudulent behaviors were employed. Furthermore, our study set a target quota of 500 respondents per country, further subdividing that quota on a per-country basis to include 250 technical managers with responsibility for introducing AI technologies into their companies and 250 operational managers tasked with using AI in their operational processes. To reach these quotas, we sampled 10,024 full-time professionals in companies with at least 100 employees, of whom 6,781 were white-collar decision-makers.

Out of these, 2688 professionals reported themselves as eligible to respond to the survey (being either decision makers or operational managers involved with AI). Out of those, only 16 people dropped out of the survey, giving an exceptionally high response rate of 93.9%. Another 147 respondents had answering patterns that were algorithmically flagged as unreliable and were removed from the sample, resulting in a net of 2525 respondents.

It should be noted that some of the 7,336 screen-outs might have been made incorrectly should these respondents have given incorrect information. Also, without knowing the composition of the population we are sampling (i.e., there are no reliable international statistics describing the population we are sampling), assessing potential systematic errors related to the selected sample is not possible.

Paper 2.

Amongst a multitude of algorithms: How distrust transfers between social and technical trust referents in the AI-driven organization

Authors:
Rebecka C. Ångström, Magnus Mähring, Martin W. Wallin, Eivor Oborn, Michael Barrett

Status:
A shorter version of this paper is accepted to the Hawaii International Conference on System Science (HICSS) -57

Abstract

Although trust has been identified as critical for successfully integrating Artificial Intelligence (AI) into organizations, we know little about trust in AI within the organizational context and even less about distrust. In this paper, we investigate how distrust in AI unfolds in the organizational setting. We draw from a longitudinal case study in which we follow a data analytics team assigned to develop numerous AI algorithms for an organization striving to become AI-driven. Using the principles of grounded theory, our research reveals that different organizational distrust dynamics shape distrust in AI. Thus, we develop three significant insights. First, we reveal that distrust in AI is situated and involves both social and technical trust referents. Second, we show that when a trust referent is rendered partly invisible to the trustor, this leads to the misattribution of distrust. Lastly, we show how distrust is transferred between social and technical trust referents. We contribute to the growing literature on integrating AI in organizations by articulating a broader and richer understanding of distrust in AI. We present a model of distrust transference actuated by social and technical trust referents. We also contribute to the literature on trust, showing how AI artifacts are implicated in trust relations within organizations.

Keywords: Artificial intelligence, organizational trust, organizational distrust, AI-driven organization, trust transference, social and technical trust referents, longitudinal case study

Introduction

As organizations launch initiatives to "become AI-driven" (Agrawal et al., 2018; Iansiti & Lakhani, 2020), they commonly introduce Artificial Intelligence (AI) to automate and transform work (Berente et al., 2021; Rai et al., 2019; von Krogh, 2018). Alongside these developments, a spectrum of concerns has come to the fore regarding the consequences of using AI technologies in and for organizations, including how AI may influence job content, job security, and human autonomy (Christin, 2017; Frey & Osborne, 2017; Kellogg et al., 2020). It is, therefore, not surprising that practitioners and scholars alike have pointed out the importance of trust for the successful integration of AI into the workplace (Candelon et al., 2021; Fountaine et al., 2019; Glikson & Woolley, 2020; Leonardi et al., 2022). Failing to establish trust in AI can result in rejection or disuse of the technology (Brayne & Christin, 2021; Dietvorst et al., 2015).

However, with few exceptions (Chawla, 2020; Leonardi et al., 2022; Lumineau et al., 2022), studies on AI and trust have focused on the direct relationship between an individual human trustor and the single AI trust referent (Jacovi et al., 2021; Lockey et al., 2021), rather than on how trust is shaped in an organizational context. This oversight is unfortunate, since organizations striving to become AI-driven typically develop many algorithms (Agrawal et al., 2018; Iansiti & Lakhani, 2020) and engage individuals and units across organizational domains (Fountaine et al., 2019; Henke et al., 2018; Iansiti & Lakhani, 2020). Furthermore, by identifying rejection as a possible outcome of interacting with AI (Dietvorst et al., 2015), the construct of distrust in relation to AI remains relatively unexplored. Drawing from the trust literature (Lewicki et al., 1998; Mayer et al., 1995; Rousseau et al., 1998), we also know that disruptive events, such as organizational transformation and technological advancement, can threaten employee trust in the organization (Dirks & de Jong, 2022; Gustafsson et al., 2021; Kähkönen et al., 2021) and lead to distrust among groups (Sørensen et al., 2011). Thus, in this paper, we take a more holistic approach to explore how distrust concerning AI evolves within an organization undertaking an effort to become AI-driven. We formulate our research question as follows: How do social and technical distrust dynamics unfold while integrating AI tools and AI-related work practices into the fabric of the organization?

We present a longitudinal case study conducted at a multinational technology firm (GlobalTech), where one business unit is undergoing a significant transformation to become AI-driven. We follow the work of a data analytics team with the assignment to develop an extensive range of algorithms, serving the frontline for realizing a corporate AI initiative, including their interaction with users. During our fieldwork, we identified different puzzling distrust phenomena concerning AI development that remained unresolved despite the developers' best efforts. Unable to explain these occurrences, we focused our empiric investigation on distrust in relation to AI.

Our findings reveal that distrust in AI is situated in and involves both social and technical trust referents. We also show that when a trust referent is rendered partly invisible to the trustor, this leads to the misattribution of distrust. Lastly, we demonstrate how distrust is transferred between social and technical trust referents. Based on these three findings, we make two key contributions. First, we contribute to the growing literature on integrating AI in organizations (e.g., Berente et al., 2021; Faraj et al., 2018; van den Broek et al., 2021) by articulating a broader and richer understanding of the crucial role of distrust in AI. We present a model of distrust transference actuated by partly invisible social and technical trust referents. Second, we contribute to the literature on trust (Fulmer & Gelfand, 2012; Lumineau et al., 2022), demonstrating how digital technology is integral in shaping trust relations within organizations.

The paper is structured as follows. First, we discuss prior research on organizations adopting AI, trust and distrust in AI, and trust in the organizational context. We then present our research approach and site, followed by the findings from our in-depth field study. Lastly, we develop and discuss our results in relation to the literature and articulate the study's contributions, limitations, and implications.

Literature

Below we define AI and discuss the literature regarding AI in organizations, trust, and distrust in AI, social influence on trust and distrust in AI, and trust transfer, which we find relevant to our examination of trust in the AI-driven organization.

What is AI?

Artificial intelligence (AI) is not a single technology but comprises several technologies, such as machine learning, natural language processing, and computer vision (Maslej et al., 2023). We follow Faraj et al. (2018) in using a single term, AI, to refer to "an emergent family of technologies that build on machine learning, computation, and statistical techniques, as well as rely on large data sets to generate responses, classifications, or dynamic predictions that resemble those of a knowledge worker" (Faraj et al., 2018, p. 62). When discussing a single AI application, we use the term 'algorithm.' Like IT, AI is a general-purpose technology (Brynjolfsson & Mitchell, 2017). However, what differentiates AI from traditional IT technology is its ability to digest vast amounts of data in order to identify patterns, predict outcomes, and propose proactive solutions (Agrawal et al., 2018; Faraj et al., 2018). Through data, AI can continue to learn and improve its accuracy (Berente et al., 2021). However, this dependency on data is also AI's vulnerability. Learning from low-quality data containing faults, biases, or missing data points will deteriorate AI reliability and potentially lead to algorithmic breakdowns (Boyd & Crawford, 2012; Danks & London, 2017; Faraj et al., 2018). Therefore, data are a vital element of AI's ability (Iansiti & Lakhani, 2020; von Krogh, 2018).

The AI-driven Organization

As a general-purpose technology, drawing from data, AI can be applied and utilized in various fields to automate tasks and processes or aid in human decision-making (Agrawal et al., 2018; Brynjolfsson & Mitchell, 2017). Transforming itself to become 'AI-driven,' an organization thus intends to develop and integrate numerous algorithms across the organization (Agrawal et al., 2018; Iansiti & Lakhani, 2020). AI is, however, not an off-the-shelf technology. Instead, building relevant algorithms requires the involvement of both domain and technical data science expertise, where the domain experts know the business processes and desired goals and outcomes (Fountaine et al., 2019; Grønsund & Aanestad, 2020; van den Broek et al., 2021). For instance, during a two-year ethnographic study, following an organization developing a system to support the hiring process of job candidates, researchers found that the developers needed to include the domain experts for selecting data, understanding the hiring process, and realizing the vision of the workforce (van den Broek et al., 2021). Because it is versatile and transformative, AI will alter jobs for employees (Faraj et al., 2018; Rai et al., 2019; von Krogh, 2018) or augment employees while performing those jobs (Agrawal et al., 2018; Rai et al., 2019). For instance, AI is expected to aid humans in performing cognitive tasks such as decision-making and problem-solving (Agrawal et al., 2018; von Krogh, 2018). This change does not come without its adverse effects, however, such as employers increasing surveillance and control over employees (Brayne & Christin, 2021; Faraj et al., 2022; Kellogg et al., 2020) or the risk of job loss (Frey & Osborne, 2017). Such adverse effects can result in human employees resisting the integration of AI (Brayne & Christin, 2021; Kellogg et al., 2020).

Trust and Distrust in AI

As AI is expected to impact organizations broadly, scholars and practitioners have pointed out the importance of trust for successful integration (Candelon et al., 2021; Fountaine et al., 2019; Glikson & Woolley, 2020; Leonardi et al., 2022). Trust is commonly defined as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau et al., 1998, p. 395). As a construct, it includes two critical dimensions: first, the trustor must have positive expectations, and second, these positive expectations are held in the presence of risk or uncertainty, calling for the trustor to willingly accept vulnerability (Fulmer & Gelfand, 2012; Rousseau et al., 1998). The roles involved in trust relations are named 'trustor', the party that is trusting, and a trust referent, the party that is trusted, also referred to as the 'trustee'. Regarding AI, the trust relationship usually includes the individual human trustor and a single algorithm as the trust referent (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021). Distrust is defined as "confident negative expectation regarding another's conduct" (Lewicki et al., 1998, p. 439), where the trustor (or distrustor) is unwilling to succumb to vulnerability (Bijlsma-Frankema et al., 2015; Lewicki et al., 1998). Distrust is non-equivalent to low trust and a separate construct from trust (Dimoka, 2010; Lewicki et al., 1998; Saunders et al., 2014). As such, it follows its own dynamic, which includes self-amplifying cycles (Bijlsma-Frankema et al., 2015). Few scholars have investigated distrust in relation to AI. However, research has demonstrated how users reject AI and develop algorithmic aversion (e.g., Christin, 2017; Dietvorst et al., 2015), a similar response to distrust whereby trustors are unwilling to succumb to vulnerability. As distrust can generate self-amplifying cycles, it is thus an important construct to explore.

Studies reveal that trust in algorithms depends on the algorithm's specific capabilities, such as transparency, reliability, and the level of task substitution (automation or augmentation) (Glikson & Woolley, 2020; Lockey et al., 2021). Violating the same capabilities can give rise to negative expectations or rejection of the algorithm. Transparency is vital for trust in AI because it provides users with the reasoning and inner logic of a specific algorithm, as it can explain why it came to its particular conclusion (Glikson & Woolley, 2020; Hoff & Bashir, 2015). More advanced algorithms can suffer from being black-boxed and incapable of explaining themselves, and even a data scientist can have difficulties revealing how the algorithm came to its conclusion (Castelvecchi, 2016). The lack of transparency can undermine trust and result in users attempting to resist or game the algorithms (Möhlmann & Zalmanson, 2017). Reliability refers to algorithms exhibiting consistent and expected behavior over time (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021; Lumineau & Schilke, 2018). When an algorithm errs, this can result in trust breakdowns, so-called algorithmic aversion, where a user becomes less confident in the algorithm and, hence, prefers human aid (Dietvorst et al., 2015). Important to notice is that the algorithms' actual reliability is not enough to warrant the user's trust; research shows that the users must also perceive them to be reliable (Lockey et al., 2021). Trust also depends on task substitution, whereby algorithms augment or automate human tasks (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021). For example, research found that algorithmic decisions were perceived as less trustworthy than human decisions when the decision involved human skills, such as hiring or evaluating work performance (Lee, 2018). Research also revealed that domain experts are more reluctant than laypeople towards algorithms augmenting or automating tasks (Hoff & Bashir, 2015; Logg et al., 2019). One reason for domain experts' skepticism could be that they perceive a risk of deskilling and loss of job security as algorithms begin to perform tasks independently and, as such, will compete with domain experts (Lockey et al., 2021).

Social Influence on Trust and Distrust in AI

Integrating AI in organizations creates a situation of uncertainty and places organizational members in vulnerable positions, as AI may threaten job content, job security, and human autonomy (Christin, 2017; Frey & Osborne, 2017; Kellogg et al., 2020). Researchers have revealed that additional actors, such as algorithmic brokers, can be relevant when establishing employee trust in algorithms (Kellogg et al., 2020; Waardenburg et al., 2022). However, trust literature tells us that other trust relations within the organizational context may be of interest, as trust relations can form intricate webs involving trustors and trust referents across levels of analysis (Fulmer & Gelfand, 2012; Lumineau & Schilke, 2018). For instance, disruptive events, such as organizational transformations and technological advancement, can threaten employee trust in the organization (Dirks & de Jong, 2022; Gustafsson et al., 2021; Kähkönen et al., 2021). Such organizational transformation can also lead to distrust between groups, whereby distrust can form self-amplifying cycles between groups (Bijlsma-Frankema et al., 2015; Sørensen et al., 2011). Few scholars have explored the social trust levels for AI (for example, Chawla, 2020; Hengstler et al., 2016; Leonardi et al., 2022; Lumineau et al., 2022). For instance, scholars have suggested that users of AI become vulnerable to developers and data providers (Lumineau et al., 2022). However, these trust referents are often described as third parties, distant and anonymous (Lumineau et al., 2022), or scholars focus on their role in establishing trust (Hengstler et al., 2016; Leonardi et al., 2022). Acknowledging the situated development of AI, engaging both technical and domain experts, and aiming to develop and integrate numerous algorithms across organizations, there are convincing reasons to believe that social relations influence both trust and distrust in AI to a greater extent than previously explained.

Trust transference

Trust and distrust can be transferred between trust referents. Trust transference is a cognitive process where "the trustor transfers trust from a known entity to an unknown one" (Doney et al., 1998, p. 605) or when a trustor bases their initial trust in one party (individual, team, or organization) on their trust in another party (Stewart, 2003). Trust transference can also occur when there are differences in the level of trust between trust referents, such as during trust repair, where the aim is to transfer trust from a party that is perceived to be trustworthy to a distrusted or discredited party (Bachmann et al., 2015; Kähkönen et al., 2021). Likewise, trust transfer can move in the other direction, where a distrusted party could damage the legitimacy of a credible party (Bachmann et al., 2015). For trust transference to occur, the trustor must be able to establish links between the parties in question (Doney et al., 1998), perceiving them to be related, for instance, by their similarity, proximity, or common faith (McEvily et al., 2003; Stewart, 2003). Research also explores trust transfer and technology where different trust referents, such as technology, providers, and platforms, influence users' trust for specific services or technologies (Belanche et al., 2014; Gong et al., 2020; Shao et al., 2022). In a recent study exploring trust transfer and AI, the only one to our knowledge, scholars revealed that survey respondents' trust in self-driving vehicles depended on their existing trust in vehicle technology, vehicle provider, and AI technology, but not AI providers, which the authors argue was due to users not yet associating AI providers with vehicle development (Renner et al., 2022). To the best of our knowledge, no research has explored the transfer of distrust in relation to AI.

Method

Our longitudinal case study focuses on a data analytics team embedded in a global technology firm. The team's assignment is to develop algorithms for a local organization as part of an initiative for the organization to become AI-driven. Following the organization for 24 months, we used multiple data collection methods, including observations and semi-structured interviews. We analyzed the data and developed a data structure using the principles of grounded theory.

Empirical Context

Our research site is located at a local operation center in Europe, part of a business unit of a multinational business-to-business (B2B) technology company (henceforth "GlobalTech"). The operation center manages geographically dispersed installations of field equipment for GlobalTech's customers. This work includes supervising the equipment, responding to equipment alarms, and dispatching and supporting field technicians serving the equipment. The local operation center manages the field equipment for approximately fifty customers; each is assigned a specific customer account and dedicated staff.

Numerous teams are involved in the operational work serving the customers' equipment. Front- and back-office teams monitor the equipment, a domain expert handles significant incidents, and customer-facing teams are physically located closer to the customers. We refer to all these teams as 'operations teams' unless it is relevant to point out their particular functions. Situated within the local organization is a data analytics team (henceforth "DA team"), the only local team with technical expertise in data science. The DA team is assigned to deliver various analytics products (i.e., algorithms) to support the operations teams. We followed the organizational transformation to implement the corporate AI strategy, specifically focusing on the occurrences among the operation teams, the DA team, and the algorithms. Furthermore, to build their algorithms, the DA team access data for their algorithms from GlobalTech systems, the customers' systems, and third parties. Generally, the data include information on equipment, alarms, tickets created for alarms, and work orders. The DA team's employees consist of five different roles (see Table 1) and, being external recruits, they lack domain expertise.

Table 1. Roles and Key Responsibilities of the DA Team

Roles	*Key Responsibilities (as described by the DA team)*
Data Analyst	Define, create, automate, and maintain key operational, performance and financial analysis. Develop analysis process to increase task effectiveness.
Data Engineer	Collect, clean, stitch together, and analyze data.
Data Scientist	*Develop best-in-class statistical models (descriptive, predictive, and prescriptive) and algorithms. Focus on machine learning, developing, and using advanced analytics methodologies to turn data into knowledge, insights, and recommendations.*
AI Business Translator	Ensure links among analytics, business, and operations teams and being able to speak each other's languages. Provide business insights to guide priorities and help identify opportunities.
Data Architects	Define standard data structures and ensure compliance, quality, and consistency of data flows. Ensure future data requirements are robust. Develop data roadmaps.

During our fieldwork, the operation center underwent a significant transformation following the business unit's new corporate strategy to become AI-driven. This strategy is embodied in a new operating model, which includes transforming the organizational structure, assignments, and roles.

Data Collection

We were granted access to GlobalTech's business unit and operation center from May 2019 to December 2021. During our fieldwork, we followed the DA team's work and witnessed several critical events, including the failure of a predictive model, a critical customer onboarding event, and the effects of the Covid-19 pandemic. In April 2021, during a reorganization, the DA team was divided into three parts. This event became the natural end to our fieldwork, which concluded with follow-up interviews with key informants between May and December 2021. During the fieldwork, we collected data both physically and digitally. We spent 16 days at the operations center, 12 of them for observations exclusively. In total, we conducted 51 semi-structured interviews and 18 recorded follow-up discussions. The fieldwork includes 31 informants that were interviewed (see Table 2), and each recorded conversation (interview or discussion) spans 30 to 120 minutes. The total recorded material is 60 hours, all of it transcribed. Furthermore, nine conversations were documented in field notes but not recorded. Regarding the data included in the paper, verbatim data were transcribed, but selected quotes were grammatically modified for readability.

Table 2. Respondents

Team	*Role*	*Recorded Interviews*	*Recorded Discussions*	*Non-recorded Meeting with Notes*	*Observation*
DA team	DA Manager 1	4	5	3	y
	DA Manager 2	5	2		y
	DA Manager 3	1			
	DA Employee 1	4	1	1	y
	DA Employee 2	4	3	1	y
	DA Employee 3	1			y
	DA Employee 4	1			
	DA Employee 5	1			
	DA Employee 6				y
	DA Employee 7	1			
	DA Employee 8				y

	DA Employee 9				y
Operation management	Operation Manager 1	3	5	1	
	Operation Manager 2	2		1	
	Operation Manager 3	1			
	Operation Manager 4	1		1	
	Operation Manager 5	1			
	Operation Manager 6	1			
	Operation Manager 7	1			
Operation Employees	Operation Employee 1	1			y
	Operation Employee 2	1			
	Operation Employee 3				y
	Operation Employee 4				y
	Operation Employee 5				y
	Operation Employee 6				y
	Operation Employee 7				y
	Operation Employee 8	1			
	Operation Employee 9	1			
	Operation Employee 10				y
	Operation Employee 11				y
Automation Employees	Automation Employee 1	1			
	Automation Employee 2	1		1	
	Automation Employee 3	1			
Customer support	Customer officer	1			
Senior Management	Senior Operation Manager	2			
	Senior Operation Consultant	1			
Corporate Management	AI Evangelist	1			
	Top Manager 1	2			y
	Top Manager 2	2			y
	Group Discussion with Top Managers 1 & 2		2		
HR	HR Partner 1	2			
	HR Partner 2	1			y
Total		**51**	**18**	**9**	

In total, 110 documents were collected, including reports, presentations, emails, internal news postings, and Yammer conversations. Due to corporate restrictions against video recording, observations were documented using field notes. These notes were transferred to Word documents immediately after the observation to minimize loss of accuracy and detail. Table 3 provides an overview of the primary data collected during the fieldwork.

Table 3. Data Collection

Year	*Type of Data*	*Frequency/Amount*
2019	Days spent at the operation center	13 days
	Time spent on observations	9 days
	Observation at GlobalTech HQ (HR workshop)	2 days
	Recorded interviews and discussions	26
	Documented discussion (field notes)	9
	Documents collected	50
2020	Observations at the operation center	3 days
	Digital observations (online meetings)	5 hours
	Recorded interviews and discussions	25
	Documents collected	36
2021	Recorded interviews and discussions	18
	Documents collected	21 (+ 3 during 2022)

Data Analysis

A thorough analysis was possible by collecting different types of data, such as semi-structured interviews, observations, and documents. Through semi-structured interviews, we followed the development of previous and ongoing critical events, collaboration, and relations as the case progressed. Collected documents helped to naturally capture issues and topics relevant to the organization and teams without probing (i.e., email, internal news postings, and Yammer conversations). (See Table 4).

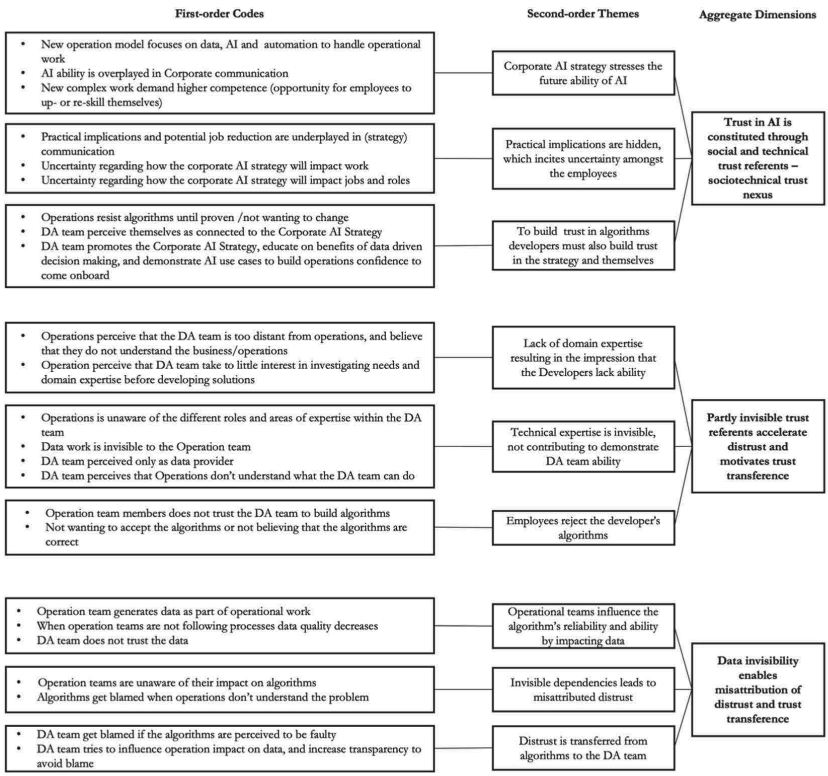
Table 4. Empirical Material

Data	*Type of Material*	*Use in Analysis*
Observations	HR workshop, May 2019 Visit at the operation center, June 2019 Observation of operational work, Aug 2019 DA team daily work, Sep 2019, Jan 2020 Operation manager meeting, Oct 2019 Digital meetings, April-May 2020	Analysis of Corporate AI strategy spoken and unspoken aims General description and understanding of work tasks and processes Studying social interactions within and between teams during collaborations
Recorded Interviews and Discussions	DA team (33) Operation Managers (15) Operation Employees (4) Automation Employee (3) Customer Officer (1) Senior Operation Managers (3) Corporate Management (7) HR (3)	Investigating expectations on AI, corporate AI strategy, and becoming AI-driven Mapping of trust development relating to the DA team and algorithms Analysis of trust relations, trust building, and trust breakdown Mapping of critical events over time and their impact on team perception Analysis of contrasting ideas and views between teams and between management levels
Documents	Reports Presentations Internal news postings Yammer conversations Emails Role descriptions Work processes descriptions	Analyzing discourse beyond interviews and observations (i.e., email and Yammer conversations) Contrasting corporate communication regarding the corporate AI strategy, organizational structure, and roles with employee perception Reconstructing the timeline and triangulation of events

We adopted the principles of constant comparisons used in grounded theory (Glaser & Strauss, 1967), where data collection and coding are conducted iteratively as the fieldwork progress. In addition, we followed the established guidelines of the Gioia Methodology for inductive concept development (Gioia et al., 2013) to ensure rigor in our qualitative research. The systematic approach includes using a data structure and coding data into first-order concepts, second-order themes, and aggregated dimensions (Gioia et al., 2013). Conducting the first-order analysis resulted in 122 codes. As distrust emerged as a central phenomenon, we identified statements and behaviors expressing positive or negative trust perceptions (Brattström et al., 2019; Lewicki et al., 1998). All trust statements and behaviors were thoroughly examined to determine the trustor, trust referent, and relevant factor of trustworthiness according to Mayer et al. (1995) and Glikson and Woolley (2020). We found four distinct trust referents: the Corporate AI strategy, the DA team, the algorithms, and the data. The Corporate AI strategy combines the organization's intentions with the transformation to become AI-driven with the employee's perception of AI in general. Though mostly invisible to the operation teams, data emerged as a trust referent throughout the fieldwork, often referred to by the DA team. Though both the DA team and the operation teams expressed mutual trust perceptions towards each other, we focused on the statements and events where the operation teams were positioned as trustors to answer the research question. In terms of the level of trust, though all teams comprise individuals, the informants often referred to "we," "us," or "them." For example, one of our informants stated, "I have nothing against DA Employee 4 as such, but they are on that island." As a result, we decided to focus exclusively on trust and distrust at the inter-team and organizational levels and not include trust at the individual level. This decision was further strengthened as analysis clarified that the DA team's different internal roles and functions were chiefly invisible to the operation teams.

Building on our initial coding, we identified themes of distrust, where encounters between both social and technical trust referents influence distrust in AI. Comparing our themes with existing concepts from the literature, such as trust in AI (Glikson & Woolley, 2020), distrust development (Bijlsma-Frankema et al., 2015; Lewicki et al., 1998; Sørensen et al., 2011), and trust transfer (e.g., Stewart, 2003), we looked for similarities and differences that could explain our phenomena. The comparison resulted in a clear set of second-order themes (Gioia et al., 2013). These themes are highlighted in cursive in the finding section's analysis paragraphs. Further distilling our second-order themes into three aggregated dimensions, we revealed three distrust dynamics that actively influence distrust during the integration of numerous algorithms. Lastly, we built our data structure by defining our first-order codes, second-order themes, and aggregated dimension (see Figure 1). For further illustrative quotes, see Table 5 in Appendix 1.

Figure 1. Data structure



Findings

Our fieldwork began in May 2019 and continued throughout 2021. This paper focuses mainly on events occurring between May 2019 and May 2020. During this period, we uncovered three distrust dynamics involving the operation teams, the DA team, and the DA team's algorithms. In May 2020, the Covid-19 pandemic occurred, which impacted the organization as an exogenous shock as the operational work had to be altered. This shift also ends the phase which is the focus of this paper. The contrasting period following the Covid-19 pandemic is described briefly in the case epilogue.

Distrust Dynamic 1: Corporate AI Strategy Triggers Fear for Job Security

On November 14, 2018, the business unit at GlobalTech announced that they have a new corporate AI strategy, including a new operating model. The corporate AI strategy focuses on AI, automation, and data. It also proclaims that the operations must be rebuilt from the ground up using AI and automation as core elements, removing monotonous and repetitive tasks from their current way of working while improving efficiency and reducing costs. The external marketing material announcing the change emphasizes AI and AI capabilities. For instance, in a promotional video, the head of the GlobalTech business unit is shown speaking with an AI that manages a customer's field equipment over a large geographical area. The AI, equipped with a natural female voice, performs the work currently assigned to the employees at the local operating center. In the video, the only task the head of the business unit must perform is to verbally accept the AI's suggested actions. It is clear that the corporate AI strategy will alter work for the employees, but what future roles and work tasks will entail is unclear to the organization. What the management has identified, however, is that the organization pyramid is 'too fat' at the bottom, meaning that there are too many low-skilled roles in the organization. The suggested solution is that employees must either upskill or reskill themselves to fit the corporate AI strategy. One of the HR partners foresees that working with AI will demand a significant change in employee competence:

> *We will move into a more intelligent environment, where people and artificial intelligence will work hand in hand in this digital space […]. I do see that, due to the kind of environment that we are moving into, there will be a change, there is going to be a huge change in the competence.* (HR Partner 2)

The communication targeting the employees focuses on the organization's reasons for changing, new tools and processes, and expectations of employees to adopt a new mindset. This information is available in internal news, workshops, and boot camps. However, as the management is unsure what job will be available after the shift, information regarding organizational structuring is kept at a high level, and practical implications regarding roles and tasks are absent. Furthermore, information regarding cost reduction and efficiency is avoided by management so as not to trigger employees to think about headcount reduction. However, the lack of clarity raises uncertainty regarding the corporate AI strategy's impact on individual roles and triggers fear for job security. Operation Manager 1 shares how his team was enthusiastic at first but that this quickly changed to concern:

> *The first time I spoke about the [corporate AI strategy] with my core team, we still only had the concept. They were quite enthusiastic. But then they started asking, "What should be my function?" "What will I do?" The concept does not say that. (Operation Manager 1)*

Furthermore, as corporate communication focuses on how AI can enable more automation, employees begin to fear for their jobs. Top Manager 1 reflects on the reaction he received during the rollout of the corporate AI strategy:

> *People recognized that no matter how well we dressed it up, their job was under threat as it stood at the time. (Top Manager 1)*

The DA team welcomes the corporate AI strategy and operating model. From the outset, they recognize themselves as part of the strategy, and they take pride in that their roles, processes, and algorithms already align with the new operating model. At the same time, they recognize that the other teams at the operation center are more reluctant towards corporate AI strategy. The DA team believes this reluctance comes from the fear of what the new operating model will entail and a lack of understanding of AI and data-driven methods. DA Manager 1 expresses this connection:

> *What is data-driven? What is a proactive approach? What is automation? People are scared about that. I think it is just because there is no clear understanding. (DA Manager 1)*

Seeing themselves and their algorithms as part of the corporate AI strategy while dependent on operation teams' accepting the new ways of working, the DA team arranges workshops and presentations to overcome the resistance. In the workshops, they demonstrate predictive algorithms and explain the value and benefits of the new way of working while promoting themselves and their algorithms.

To summarize this trust dynamic: The corporate AI strategy stresses the future ability of AI and overplays AI capabilities while pushing for a shift in employee skills. Simultaneously, practical implications are hidden, which incites uncertainty and vulnerability among the employees at the local organization as they begin to fear for their job security. This fear results in distrust of the corporate AI strategy. Having high trust and expectations in the corporate AI strategy, the DA team actively associates themselves with the corporate AI strategy. This association, however, enables distrust transfer between the corporate AI strategy and the DA team. Therefore, to successfully build trust in their algorithms, the DA team must overcome its distrust of the corporate AI strategy.

Distrust Dynamic 2: Perception of the DA Team Leads to the Rejection of Algorithms

When the DA team approaches the operation teams with their algorithms, the operation teams find that the DA team is too distant from the operation to understand their needs. They complain that the DA team does not understand the operation and that their algorithms are of no value. The operation teams even start to question whether the DA team is interested in learning about the operation teams' needs. For example, a customer officer argues that the DA team is not even interested in talking to the people with the domain expertise to learn what the domain experts want them to address:

> *They [the DA team] are just doing things on their own, without understanding what they are doing […] they are just developing something. They think that it is the way forward, and they are not listening to the, you know, the real experts. (Customer Officer)*

At the same time, the operation teams lack an understanding of the DA team's work and technical expertise. For example, most of the DA team's work includes different forms of data handling and algorithm preparations, which are invisible to the operation teams. DA Manager 2 explains that he believes only a small part of the team is visible to the operation teams:

> *It is hard for me to explain every time the work of a data scientist and what people are involved in building a model. Besides the fact that they see a visualization there, there are people doing data engineering, data*

> *modeling, and data understanding. In fact, they are seeing only one person, the data scientist. (DA Manager 2)*

The perceived lack of domain expertise, and invisible technical expertise, leads to the operation teams' rejection of the DA team's algorithms. For instance, during one of our stays at the operation center, we spent time with a team of domain experts. While learning about their work, we asked about their collaboration with the DA team. Two of the operation employees we spoke to express skepticism towards the DA team and explain that their expectations of the DA team are low:

> *Fieldnote: They both seem skeptical. The [Operation Employee 4] explains that they are not interested in collaborating with [the DA team]. "We manage with our own resources." They laugh and tell me that when [the DA team] deliver something, it goes wrong." (Excerpt from fieldnotes)*

The resistance towards the DA team's algorithms becomes visible in a particular case. A data scientist from the DA team develops an algorithm that predicts when field equipment is at risk of malfunctioning due to hot weather. The possibility of predicting overheating equipment would allow the operational team to take preventive actions; for instance, sending field technicians to cool down the equipment. However, the operational team managing the customer account is not interested. To convince them of the algorithm's value, the data scientist emails the team every time the algorithm predicts that a piece of equipment is at risk of overheating. She sends the emails for almost a year before the team accepts the algorithm. The data scientist believes the operation teams' resistance is due to distrust of the algorithm. However, when we asked the Operation Manager 1 about the same case, he said he believed there was a lack of trust towards the data scientist, as she had not taken the time to investigate the operation needs:

> *Interviewer: was there a lack of trust in the [algorithm] or the data scientists?*
>
> *CO Director: I think it was a lack of trust for the data scientist. (Operation Manager 1)*

To summarize this trust dynamic: During collaborations, the operation teams perceive that the DA team lacks domain expertise, resulting in the impression that it lacks ability to build algorithms of value. At the same time, the operation teams do not understand the work that the DA team performs nor the constraints that bind it. As such, the DA team's technical expertise is invisible, accelerating distrust in the team, as it hides their actual ability. This distrust results in the operation teams distancing themselves from the DA team. Lastly, the distrust transferred to the algorithms is manifested in the operation teams' rejecting the DA team's algorithms.

Distrust Dynamic 3: Data Issues Create Distrust in Algorithms and are Blamed on the DA Team

The operation center's domestic IT tools are a significant data source for the DA team. While managing the field equipment, operational employees interact with different IT tools to fill in information and respond to system output. The information is used to track the incident occurring with the field equipment, analyze the root cause of the incidents, and communicate with other teams and customers. Their actions and inactions generate data that the DA team extracts for their analysis and algorithms. If the operation employee is not following the operational processes, not adding standardized information, or missing adding information altogether, this impacts the quality of the data gathered from the tools. This decrease in data quality in turn impacts the DA team's algorithms' reliability. For example, during the summer of 2019, one of the DA team's predictive algorithms inaccurately overestimated how long a piece of field equipment could manage without service, resulting in an equipment failure. At first, the operation teams believed that an operation employee had made a mistake, but as they investigated the issue, it became clear that the algorithm's prediction was wrong. The DA team, however, conducted a subsequent investigation revealing that the algorithm had learned from data containing an operation employee mistake, which had been included in the algorithms training data. The DA Manager 1 comments on the human error leading to the faulty prediction:

> *We need a little bit of discipline in our work. We need to understand that if we are not disciplined with what we are doing, and we don't believe in our data, we cannot become data-driven.* (DA Manager 1)

Data access is also a challenge for the DA team when developing algorithms. During one of our stays, we witnessed how customer data access is challenging for the DA team. In January 2020, one of GlobalTech's call centers performed poorly, and the DA team was asked to analyze the cause. The issue was a high priority that engaged operational teams, managers, and senior managers. The DA team quickly realized they needed data from the customer's system to provide an effective analysis. However, accessing customer data must be handled by the operation teams, and this is not a prioritized task for the operation teams. Hence, instead of gaining access to the system directly, the DA team receives daily data dumps in Excel files sent in emails from the customer. There was little consistency among these daily files, and thus the DA team struggled to form a coherent analysis of the call center's performance. As a result, operation teams and managers rejected the DA team's analysis regarding the call center, as they perceived the report as being deficient. The DA Manager 1 reflects on how she believes that scarcity of data is the source of the rejection:

> *We are receiving by email some snapshots [excel sheets of data]. We are running our analysis based on those snapshots. The fact that those snapshots are not complete is not our fault. And they say, "No, the report is not good." So, there is a huge resistance. (DA Manager 1)*

One of the operation center's senior operation managers recalls the customer onboarding events a year later and that the DA team claimed they lacked data access. However, the Senior Operation Manager sees it as a poor excuse and finds that the DA team has become too comfortable:

> *When things are really bad, I don't care if someone needs to have a clipboard and walk around and ask people in the [operation center] how many tickets did you create today. (Senior Operation Manager)*

The operation teams are unaware of how their work impacts the data quality and access and how this, in turn, impacts the DA team's algorithms. So instead, they blame the failing algorithms on the DA team and start questioning the DA team's ability to build reliable algorithms. A member of the DA team reflects on how the blame shifts:

> *I mean, they are blaming the model for the problems. But the problems are not because of the model but because of the data accuracy behind the model. They are blaming the team that they didn't do a good model, but the problem, in fact, stays in the data, and this is what we tried all the time to explain: "please understand, garbage in, garbage out." Yeah, so*

> *if the data are inaccurate, don't blame the model. Yeah, don't blame the team that build the model. (DA Employee 1)*

To summarize this trust dynamic: The operational teams influence the algorithm's reliability and ability by impacting data. They do this in two ways. First, if the operation employees neglect standard processes when handling the operation tools, this deteriorates data quality and decreases algorithm reliability. Second, the operation teams control part of the data access, constraining the DA team's development of algorithms, which impacts the algorithm's ability. In both cases, the algorithms' invisible dependencies lead to misattributed distrust, whereby operation teams assign the fault to the algorithms. Their distrust is transferred from the algorithms to the DA team.

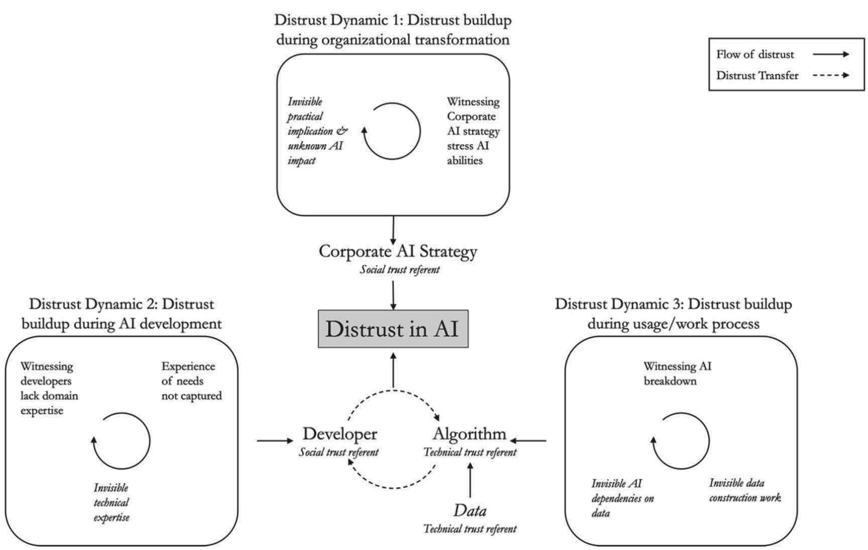
Case Epilogue

The challenging situation of developing and integrating algorithms into the organization continues to unfold until the pandemic of Covid-19 starts to spread across Europe. The pandemic becomes an exogenous shock to the organization as GlobalTech issuing new working directives forces employees to work from home. Removed from the physical environment, the organization finds itself dependent on data analytics to run the operational work, increasing the requests for the DA team's algorithms. The development starts to be conducted together with a domain expert team. At the end of 2020, the DA team developed 50 unique algorithms implemented across customer accounts. The increased experience with algorithms builds operation teams' trust in algorithms. However, the perception of the DA team's lack of domain knowledge is still strongly manifested in the organization, and the team is not trusted to deliver on new requests. The trust in algorithms is, as such, not transferred from algorithms to the DA team. Instead, it transfers to the domain expert team. In January 2021, the Senior Operation Manager announced a reorganization where the DA team was split into three parts. Senior Operation Manager's main reason for the restructure is the slow uptake of algorithms, and the DA team has not been able to integrate properly into the organization. In the new organization, algorithmic development is entrusted to the Operation Manager 2, the manager of the domain expert team who became involved in the development of algorithms during the Covid-19 pandemic.

Discussion

Our study addresses how distrust in AI unfolds in the organizational setting. On the bases of our in-depth field study, we develop a model of distrust transference during the development and integration of AI tools and AI-related work practices into the organization (see Figure 2). Our model describes how partly invisible trust referents within an organization result in distrust dynamics where distrust transfers between trust referents while establishing trust in AI.

Figure 2. Model of Distrust Dynamics Related to AI in a Multi-Algorithmic Organizational Setting



We began this paper with a simple observation. Prior research has argued that failing to establish trust in AI can result in rejection or disuse of the technology (Brayne & Christin, 2021; Dietvorst et al., 2015). Rallying around such consensus, a large body of research has emerged that examines AI and trust between the individual human trustor and a single algorithm trust referent (Jacovi et al., 2021; Lockey et al., 2021). However, this perspective is severely limiting for two reasons. First, the construct of distrust is seldom explored as part of the algorithmic rejections. Second, the efforts to integrate AI are often part of corporate-wide initiatives involving a numerous of algorithms and spanning individuals, teams, and units, likely creating an intricate web of relationships (Fountaine et al., 2019; Henke, Levine, & McInerney, 2018; Renner et al., 2022). To address this gap, we set out to investigate how social and technical distrust dynamics unfold when integrating AI into organizations. We developed a range of explanations rooted in a socio-technical (Mumford, 2006) understanding of trust in AI. First, we identify that distrust in AI involves both social and technical trust referents. Second, we recognize that distrust emerges when trust referents are not rendered completely visible to the trustor. Third, we demonstrate that distrust is transferred between the social and technical trust referents.

Distrust in AI Depends on Both Social and Technical Trust Referents

Our first finding places AI in a situated context, revealing how both social and technical actors trigger distrust in AI that emerges, forms, and blends into the organization. We identify two social trust referents, the 'corporate AI strategy' and the 'developers,' and two technical trust referents, the 'algorithms' and 'data.'

Research in organizational trust has revealed that a corporate strategy containing operational and human resources strategies can affect employees' perception of the organization's trustworthiness (Gillespie & Dietz, 2009). Furthermore, employees' trust in organizations and managers can become challenged during major organizational transformations (Gustafsson et al., 2021; Sørensen et al., 2011). Our research reveals a similar pattern where the organization's decision to become AI-driven results in uncertainty among employees. However, we also see that introducing AI adds a new dimension of uncertainty based on the unknown potential of AI to overtake tasks and job roles. This uncertainty grows as corporate communication portrays AI as futuristic while practical implications for job impact are absent in the corporate AI strategy. We refer to this development as Distrust Dynamic 1, illustrated in our model. The result is a distrust in the corporate AI strategy, which includes the organization's intention and the AI potential. Hence, we conclude that a corporate AI strategy can (and is likely to) function as a trust referent.

Our second identified trust referent is the developers. Scholars have begun to recognize that trust in developers plays a part in shaping users' perception and adoption of AI (Chawla, 2020; Leonardi et al., 2022; Lumineau et al., 2022). However, current research does not explain how trust and distrust for AI are influenced by developers situated within the organization and continuously collaborating with surrounding teams on new algorithms. We identify the developers as trust referents at the team level within the organization (Fulmer & Gelfand, 2012). Drawing from the trust literature, the experience of trusting a party will influence the trustor's perception of a trust referent for future occasions (Mayer et al., 1995). This is relevant during the development of numerous algorithms, as relations between domain experts and technical experts are continuous.

Our technical trust referents, the algorithms and the data, are related since algorithms depend on data to learn and undertake tasks (Faraj et al., 2018). We know from existing research that the algorithms' trustworthiness depends on algorithms' capabilities (Glikson & Woolley, 2020; Lockey et al., 2021). For instance, experiencing an algorithm having a reliability breakdown can create algorithmic aversion (Dietvorst et al., 2015). However, research seldom goes on to explore what causes such breakdowns (Glikson & Woolley, 2020). By separating the algorithms from the data, our research reveals that they are different trust referents and that trust, or distrust, in one of them does not necessarily mean there is trust or distrust for the other. We also identify that the perception of the algorithm's trustworthiness is dependent on data, which, to our knowledge, is seldom discussed in the literature on trust or distrust in AI (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lockey et al., 2021).

Invisibility of Trust Referents Results in Misattributed Distrust

Our second insight reveals that when trust referents are less than fully visible to trustors, it can result in the misattribution of distrust. A particularly interesting aspect of this is the invisibility of data. Scholars have pointed out that data curation can be invisible (Sachs, 2020; Waardenburg et al., 2022) and that data curation work can be performed by invisible workers (Kellogg et al., 2020). Data construction can also be part of employees' daily work (Waardenburg et al., 2022). We expand on this research by identifying how data and related data work can be invisible also to the people performing it. The invisibility of data results in algorithms becoming partly invisible too, which has consequences for AI. In our research, we identify two ways in which this plays out. First, when employees do not see how their work impact data quality, they do not know that they contribute to algorithmic breakdowns. Instead, they blame the algorithms for poor reliability. Second, when employees do not see how their work constrains data access, impacting algorithms, they do not challenge data scarcity. Instead, they blame the algorithms for lack of ability. The challenge of invisibility also affects the developers as trust referents. When employees notice the developers lack domain expertise but fail to see their technical expertise, they do not understand the developers' choice due to technical constraints. Instead, they blame the developer's ability to develop algorithms. The challenges with invisible data and data work are illustrated in Distrust Dynamics 2 and 3. The misattribution of distrust to algorithms and developers further supports our argument that trust and distrust in AI must be studied beyond the individual relations between a human trustor and the AI trust referent (Glikson & Woolley, 2020; Lockey et al., 2021).

Dependent Trustworthiness Enables Distrust Transfer

Distrust transference occurs, as noted, when a trustor's trust or distrust towards one trust referent transfers to another trust referent (Bachmann et al., 2015; Doney et al., 1998) which can occur when trust referents are perceived as related, for instance by their similarity and proximity (McEvily et al., 2003; Stewart, 2003). Our study shows that the relatedness between developers and algorithms enables distrust to transfer between the two, as illustrated in our model. Our study also shows that the relatedness between the developers and the corporate AI strategy forced the developers to counteract distrust in the corporate AI strategy, avoiding transferring the distrust to themselves and their algorithms. As such, our research reveals that distrust transfer can occur between social and technical trust referents within the organization. Furthermore, in contrast to previous research exploring distrust cycles (Bijlsma-Frankema et al., 2015; Sørensen et al., 2011), we show that distrust persists despite the best efforts of the developers to overcome it. What enables the distrust cycle to continue, is the co-dependency between the developers and the algorithms trustworthiness. For instance, when trustors perceive that the developers' lack ability, they transfer the distrust to the algorithms the developers produce by questioning the algorithm's ability. Likewise, when the trustors perceive that algorithms have limitations in ability or issues with reliability, they transfer the distrust to the developers and question their ability.

Distrust Cycles Stall the Transformation Process

To conclude, studying how trust in AI develops over time, our insights differ from findings by Glikson and Woolley (2020) showing that initial trust for embedded algorithms is high but slowly deteriorates over time. Similar to Christin (2017), who studied the introduction of algorithms in journalism and criminal justice, and Kellogg et al. (2020), who argue that resistance is a strategy to avoid algorithmic control, we find that algorithms are met with resistance from employees. However, we expand on these insights by connecting resistance with distrust, influenced by both social and technical trust referents. We identify the source of distrust as fear of job security and partial invisibility of trust referents, which challenge the work's status quo. Furthermore, we reveal that distrust transfers, enabling distrust cycles to occur, which in turn keeps the trust trajectory low over time. This distrust cycle also stalls the organizational transformation process as the uptake of algorithms is slowed. Lastly, we also expand on algorithmic aversion, the idea that seeing an algorithm perform, especially err, lowers trust in algorithms and results in a preference for human aid (Dietvorst et al., 2015). While we agree that a trustor seeing an algorithm err will likely decrease trust in the algorithm, we further argue that resistance to algorithms is a far more complex phenomenon than can be explained by the algorithm's reliability issues alone.

Conclusion

We know that the integration of AI in organizations will be far-reaching. It will be applied in all parts of the organization and for various purposes (Brynjolfsson & Mitchell, 2017), where a multitude of algorithms will be granularly built into our tasks and processes (Agrawal et al., 2018; Berente et al., 2021; Iansiti & Lakhani, 2020). Furthermore, AI's unique capabilities to mimic human intelligence, aiding us in decision making (Agrawal et al., 2018; von Krogh, 2018), will allow the technology to become woven into the social fabric of organizations. Realizing the potential impact of AI, we must continue to push forward to create a better understanding of how human jobs, autonomy, and roles and relations are altered in the AI-infused organization (Christin, 2017; Frey & Osborne, 2017; Kellogg et al., 2020; Waardenburg et al., 2022).

Such exploration will demand a socio-technical perspective and an understanding of key factors such as trust and distrust (Fulmer & Gelfand, 2012; Lewicki et al., 1998). Our study reveals that in the organizational context, by continuously introducing a multitude of algorithms, borders between social and technical domains are increasingly fluid and distrust can form an intricate web between social and technical trust referents. We contribute to the IS literature by demonstrating how distrust can spiral when organizations strive to become AI-driven. We do this by presenting a model for distrust transference between social and technical trust referents in the organization. We also contribute to the trust literature by establishing how digital technology shapes distrust relations within organizations.

References

Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: The simple economics of artificial intelligence. *Harvard Business Review Press.*

Bachmann, R., Gillespie, N., & Priem, R. (2015). Repairing trust in organizations and onstitutions: Toward a conceptual framework. *Organization Studies*, 36(9), 1123–1142.

Belanche, D., Casaló, L. V., Flavián, C., & Schepers, J. (2014). Trust transfer in the continued usage of public e-services. *Information & Management*, 51(6), 627–640.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450.

Bijlsma-Frankema, K., Sitkin, S. B., & Weibel, A. (2015). Distrust in the balance: The emergence and development of intergroup distrust in a court of law. *Organization Science*, 26(4), 1018–1039.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679.

Brattström, A., Faems, D., & Mähring, M. (2019). From trust convergence to trust divergence: Trust development in conflictual interorganizational relationships. *Organization Studies*, 40(11), 1685–1711.

Brayne, S., & Christin, A. (2021). Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems*, 68(3), 608–624.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, 358(6370), 1530–1534.

Candelon, F., Charme di Carlo, R., & Mills, S. D. (2021, October 7). AI-at-scale hinges on gaining a "social license." *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/ai-at-scale-hinges-on-gaining-a-social-license/

Castelvecchi, D. (2016). Can we open the black box of AI? Nature, 538(7623), 20–23.

Chawla, C. (2020). Trust in blockchains: Algorithmic and organizational. *Journal of Business Venturing Insights*, 14, e00203.

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, 4(2), 205395171771885.

Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. International Joint Conferences on Artificial Intelligence (IJCAI 2017), 4691–4697.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology*: *General*, 144(1), 114–126.

Dimoka. (2010). What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. *MIS Quarterly*, 34(2), 373.

Dirks, K. T., & de Jong, B. (2022). Trust within the workplace: A review of two waves of research and a glimpse of the third. *Annual Review of Organizational Psychology and Organizational Behavior*, 9(1), 247–276.

Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. *The Academy of Management Review*, 23(3), 601-620.

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70.

Faraj, S., Renno, W., & Bhardwaj, A. (2022). AI and uncertainty in organizing. In M. A. Griffin & G. Grote (Eds.), *The Oxford handbook of uncertainty management in work organizations* (1st ed., pp. C4.S1-C4.S15). Oxford University Press.

Fountaine, T., McCarthy, B., & Saleh, T. (2019, July-August). Building the AI-powered organization. *Harvard Business Review*, 97(4), 62-73. https://hbr.org/2019/07/building-the-ai-powered-organization

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, 38(4), 1167–1230.

Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, 34(1), 127–145.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia Methodology. *Organizational Research Methods*, 16(1), 15–31.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. Aldine.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.

Gong, X., Zhang, K. Z. K., Chen, C., Cheung, C. M. K., & Lee, M. K. O. (2020). What drives trust transfer from web to mobile payment services? The dual effects of perceived entitativity. *Information & Management*, 57(7), 103250.

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614.

Gustafsson, S., Gillespie, N., Searle, R., Hope Hailey, V., & Dietz, G. (2021). Preserving organizational trust during disruption. *Organization Studies*, 42(9), 1409–1433.

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust: The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105–120.

Henke, N., Levine, J., & McInerney, P. (2018, May 2). You don't have to be a data scientist to fill this must-have analytics role. *Harvard Business Review*, Digital publication, 8. https://hbr.org/2018/02/you-dont-have-to-be-a-data-scientist-to-fill-this-must-have-analytics-role#comment-section

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434.

Iansiti, M., & Lakhani, K. R. (2020). Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. *Harvard Business Review Press*.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 624–635.

Kähkönen, T., Blomqvist, K., Gillespie, N., & Vanhala, M. (2021). Employee trust repair: A systematic review of 20 years of empirical research and future research directions. *Journal of Business Research*, 130, 98–109.

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 205395171875668.

Leonardi, P. M., Barley, W. C., & Woo, D. (2022). Why should I trust your model? How to successfully enroll digital models for innovation. *Innovation*, 24(1), 47–64.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, 23(3), 438–458.

Lockey, S., Gillespie, N., Holm, D., & Asadi Someh, I. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. Proceedings of the 54th Hawaii International Conference on System Science.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Lumineau, F., Schilke, O., & Wang, W. (2022). Organizational trust in the age of the fourth industrial revolution: Shifts in the form, production, and targets of trust. *Journal of Management Inquiry*, forthcoming.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.

McEvily, B., Perrone, V., & Zaheer, A. (2003). Trust as an organizing principle. *Organization Science*, 14(1), 91–103.

Möhlmann, M., & Zalmanson, L. (2017). Hand on the wheel: Navigating algorithmic management and Uber drivers'. In Autonomy', in proceedings of the international conference on information systems (ICIS), Seoul South Korea (pp. 10-13).

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal*, 16(4), 317–342.

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Towards human AI hybrids. *MIS Quarterly*, 43(1), iii–ix.

Renner, M., Lins, S., Söllner, M., Thiebes, S., & Sunyaev, A. (2022, January 4). Understanding the necessary conditions of multi-source trust transfer in artificial intelligence. Proceedings of the 55th Hawaii International Conference on System Sciences.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.

Sachs, S. E. (2020). The algorithm at work? Explanation and repair in the enactment of similarity in art data. *Information, Communication & Society*, 23(11), 1689–1705.

Saunders, M. N., Dietz, G., & Thornhill, A. (2014). Trust and distrust: Polar opposites, or independent but co-existing? *Human Relations*, 67(6), 639–665.

Shao, Z., Zhang, L., Brown, S. A., & Zhao, T. (2022). Understanding users' trust transfer mechanism in a blockchain-enabled platform: A mixed methods study. *Decision Support Systems*, 155, 113716.

Sørensen, O. H., Hasle, P., & Pejtersen, J. H. (2011). Trust relations in management of change. *Scandinavian Journal of Management*, 27(4), 405–417.

Stewart, K. J. (2003). Trust transfer on the world wide web. *Organization Science*, 14(1), 5–17.

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly*, 45(3), 1557–1580.

von Krogh, G. (2018). Artificial intelligence in organizations: New opportunities for phenomenon-based theorizing. *Academy of Management Discoveries*, 4(4), 404–409.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science*, 33(1), 59–82.

Waardenburg, L., Sergeeva, A., & Huysman, M. (2022). Juggling street work and data work: An ethnography of policing and reporting practices. *Academy of Management Proceedings*, 2022(1), 16697.

Appendix 1. Illustrative Quotes from the Data

Table A1. Illustrative Quotes from the Data

Aggregated dimensions	*Second order theme*	*First-order codes and illustrative quotes*
Trust in AI is constituted through social and technical trust referents – sociotechnical trust nexus	Corporate AI strategy stresses the future ability of AI	**New operation model focuses on data, AI, and automation to handle operational work.** - It fundamentally changes our way or operating [field technology] from reactive to proactive, leveraging data, automation, and artificial intelligence (Head of Business Unit, quote in the press release) -[Corporate AI Strategy] is a methodology that will help us survive, that will help us sustain in coming years. and we will be using automations, we do have automations at the moment, but they are highly human assisted automations. (Operation Manager 6) **AI ability is overplayed in Corporate Communication** -New AI based [Corporate AI strategy] makes [operational work] simple (press release) - "Leveraging AI and Natural Language Processing, the [corporate AI strategy] automates service desk tasks, reducing reactions time and alleviating service providers from simple recurring tasks" (Corporate AI strategy presentation) - "This is history in the making" (Executive, Internal news positing) **New complex work demand higher competence (opportunity for employees to up/re-skill themselves)** - People need new skills combining data science and network engineering (Bootcamp presentation) -You would need people that monitor the machine, ok, to make sure that it is running. That in itself is a higher competence than somebody taking an alarm, taking a ticket, creating a work order, you know, pushing something along (HR Partner 1)

Aggregated dimensions	*Second order theme*	*First-order codes and illustrative quotes*
	Practical implications are hidden, which incites uncertainty amongst the employees	**Practical implications and potential job reduction are underplayed in (strategy) communication**. - I deliberately never use the word efficiency for the reason, I program myself not to use efficiency in the [Corporate AI Strategy], though it is a big part of it, because that's, that's a familiar trigger to everyone here. Everyone who deals in operations, that as soon as you mention efficiency, it's about headcount reduction. (Senior Operation Manager) **Uncertainty regarding how the corporate AI strategy will impact work** - We have those pretty slides that explain what it should look like in theory […] And they don't necessarily see the dots between what the pretty pictures, what we are selling, and how it will impact their actual activities (Operation Manager 5) - Because people are afraid, people are afraid that you will see that Rebecca is opening 10 tickets and Patricia is opening only two, that Rebecca is coming at 8:00 o'clock in the morning, for 9:00 o'clock and Patricia is coming at 10:00 o'clock. and so on and so forth. So, people are really afraid (DA Manager 1) **Uncertainty regarding how the corporate AI strategy will impact jobs and roles** - OK, people are quite reluctant or afraid about [Corporate AI Strategy], because they consider that they can lose their jobs (Automation Employee 1) - I start by asking her about the [Corporate AI Strategy]. She lets me know that she sees it as a concept, but that is has made people uncertain about their job, e.g. Will I still have my job? What if I want to buy a house? What will my job be in the future? (Discussion with Operation Manager 4)
	To build trust in algorithms developers must also build trust in the strategy and themselves	**Operations resist algorithms until proven** - Till you will be able to see some models working, some conclusions, some real facts, everyone will be resisting (Operation Manager 1)

Aggregated dimensions	*Second order theme*	*First-order codes and illustrative quotes*
		- It's a question of, how can I say, commodity, fright in some cases, and reluctancy to change. I don't know. It's a mission. So, people simply want traditional way of working. They don't feel comfortable, they don't feel open to change (DA Manager 1)
		DA team perceive themselves as connected to the Corporate AI Strategy
		- We are [Corporate AI Strategy] Compliant. One of the very few, at the moment, departments that is actually already on board with the new processes (DA Employee 2)
		-I think [DA team]is one of the most advanced [teams] into the Corporate AI Strategy. (DA Manager 1)
		DA team promotes the Corporate AI Strategy, educate on benefits of data driven decision making, and demonstrate AI use cases to build operations confidence to come onboard
		- By demoing predictive models, and also by having a lot of focus groups that are oriented on more predictive models and their benefits, and with these types of models to support operation and problem management, and so on, to become more proactive. And I think this is the step. And the more we do that, the more we have use cases and proof of concept and so on that is already proving to give results, the more confident other teams will become and the more open to come on board. (DA Employee 2)
		- Part of the [AI business Translator role], I mean, my role, how I feel it, is also to promote the new approach and to try to explain as much as possible, and to bring on board. (DA Employee 2,)
		- We consider [educating people] as duty. So, whenever we can, we try to perform some workshops. For instance, we try to do some workshops with [operation teams] to help them understand how they can work with [algorithms] (DA Manager 2)
Partly invisible trust referents accelerate distrust and motivates trust transference	Lack of domain expertise resulting in the impression that the developers lack ability	**Operations perceive that the DA team is too distant from operations, and believe that they do not understand the business/operations**
		- [DA team] were too far off operations, or the reality. Again, reality on the ground (Operation manager 1)

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
		- I'm saying to the guys "sorry, that just doesn't sound right" and everyone says "no, no, no that is correct", "I mean, you are telling me our average time to fix the most important faults we have is four and a half thousand hour, 4500 hours?". […] I said to DA Manager 1 then "we need to start operationalizing the team a little bit. I don't want [Operation center] engineers, I don't want back-office engineers, but I need people that are going to think 4 ½ thousand hours for a high priority fault average is wrong, it is something not right here." (Senior Operation Manager)
		Operation perceive that DA team take to little interest in investigating needs and domain expertise before developing solutions
		- Well, I have nothing against [DA Employee 4] as such, but they are on that island. And I don't see his involvement, or his, he is not really into the operations. He's on this island, you know. I don't see this added value to honest. Pure practically, what happens is he looks into Cognos, and he sees, oh there is a peak, or there is an increasing trend, he takes snapshot and he send it to the problem manager. (Customer Officer)
		- I saw a lot of reports [from DA team] which was not necessarily useful or were redundant, as in saying something that we would already know, knew, and could not use it (Operation Manager 3)
	Technical expertise is invisible, not contributing to demonstrate DA team ability	**Operations is unaware of the different roles and areas of expertise within the DA team**
		- The Data Engineers, the Data Architects, the Business Translators, they are all the same as Data Scientist in Operation Manager 1's eyes. He is unaware of their different roles, task and assignments (Field Observation/discussion with Operation manager 1)
		- I don't think that there is a clear understanding of what a data scientist really means (DA Manager 1)
		Data work is invisible to the Operation team
		- Because people yet don't understand what [DA team] is doing and what is the result of it (DA Manager 2)

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
		-Everybody perceived like AI/ML use cases to be like a software. There is not a good or a correct understanding of the difference between lifecycle management for ML, and a software.(DA Manager 1, discussion on algorithms continuous need for data and adjustment)
		DA team perceived only as data provider
		- [DA team] are the data authority, and I think they should bring us about what ever data we need. (Operation Manager 2)
		- Many times the [Operation teams] just ask for the raw data. they shouldn't ask for the data, because we're not data providers. They should ask for the analysis for insight they need (DA employee 3)
		DA team perceives that Operations don't understand what the DA team can do
		- "I want a total calls", "OK, what do you mean by total calls?". I mean they don't know exactly what they want and because of that we need to help them to set the perspective of the data, and how the data are, and what is the benefit of asking something instead of something else. (DA Employee 1)
		- We're dependent on the business and sometimes they don't even know what to requests (DA Employee 2)
	Employees reject the developers' algorithms	**Not trusting he DA teams to build algorithms**
		- At first, we were looked at "Why do you come to show me this and recommend me what to do?". Because that was the first reaction also from the [Operation team]. OK, we are not proficient with technical stuff in the [equipment], but we can show them why this trend of alarms increased. (DA Manager 2)
		- "What do you mean you're going to bring in some sort of a model that's going to tell us what's happening here?" And that's going to be, in effect I suppose, downplaying [The DA team's] role. I think they're going to look at it in terms of that type of resistance as well (Top Manager 1)
		- It was very funny, because we moved the report, some of the reports for one customer into Tableau. And it was easier, it was nicer. You can

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
		do it whenever you want, you can do it from your phone, you can do it from… the Delivery manager was calling me "Why you did this to me? I said "what, what I did to you?", "Why, why, it's so complicated. I don't want to enter and click". I remember that first I was shocked because the person considered that my guys had something against him. (DA Manager 1) **Not wanting to accept the algorithms or not believing that the algorithms are correct** - The first reaction is "not to be trusted". Because it's something new, they [Operation teams] don't know exactly how it is built (DA Manager 2) - What I see is a huge resistance. […]. Because whenever people saw something that is not what they want to see, they say report is not good. (DA Manager 1) - There were also people being accustomed to do things in a specific way or using specific tools. Then the new tool that was in place was working probably not as expected, and it was easy for them to say "OK, I don't want work with this one, because it's not yet in the best shape, so I'm still using the old one to raise a ticket" (DA Manager 2) - The resistance in adoption it's, it's huge. […], we don't want to adopt, we are still with Excel files. (DA Manager 1)
Data invisibility enables misattribution of distrust and trust transference	Operational teams influence the algorithm's reliability and ability by impacting data.	**Operation team generates data as part of operational work** - [Operation Employee 11] notice that an alarm has sorted itself out. That no engineers were dispatched, still the alarm stopped. He is surprised and notes this down in the ticket (fieldnotes, observing the work of Operation Employee 11) - [Operation Employee 7] tells me that sometimes logs are missed, for instance in stressful situations where the IM can be in a customer call and miss to add the information. This happens. But [Operation Employee 7] continues, the logs are all checked by [Operational Employee 6] before sent to the client or customer and if she finds that there are questions regarding the timestamps she asks. (Fieldnotes, discussion with Operational Manager 7)

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
		When operation teams are not following processes data quality decreases - If they are not using correct the tools, they're not using correct the process, the data is affected. And we will report something which is probably not the reality. I mean we need to link this. Data is an abstract, Ok, so, it's something which is follow somebody's action. If that particular person didn't follow the process, didn't follow the rules on the tools, like that, then the data is just an abstract number (DA Employee 1) - If they are not filling in correctly the node, on which the ticket is appearing, how can analytics do miracles to guess the node (DA Manager 1) - If one engineer for example forget to put a root cause or something, or he put it in his own words, not in the standard way to out it, so that particular information will be lost (DA Employee 1) -How to be a data driven company when the main assets, which is data, is not available or it's a very poor quality? And I need to spend half of my time cleaning the data or understanding why it's not enough, and we want to be data driven (DA Employee 7)
		DA team is not trusting the data -If the data is not OK, and if the data is not available, we cannot do any analysis and we cannot be data-driven. (DA Employee 1) - We need to, respect if you want, our data. Because imagine everything that we're doing is influencing the data quality. The fact that I'm not following the process, the fact that I don't follow the process impact the data. The fact that I don't have access to the right data, yeah, I don't have access to the right database. Everything is impacting our data, so I don't trust data. (DA Manager 1) - So, in six months we had so many data availability issues that I needed to keep a separate track. At least once in every two weeks we have issue with data delivery, with data availability (DA Employee 7)

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
	Invisible dependencies between operation work, data and algorithms leads to misattributed blame	**Operation teams are unaware of their impact on algorithms** - They are seeing only the end of the process. The final data. But they're not taking care of what is until the data is on the screen (DA Employee 1) - I have the feeling that they don't even know how much important work they are doing, during the flow. Yeah, people are not aware of their importance (DA Employee 1) - "There is no need to go back to the guys and ask for putting a correct root cause or filling some fields in a proper way. You should do miracles with data" (DA Manager 1 repeats a statement from the operation team) **Algorithms get blamed when operations don't understand the problem** -The model was blamed, but the model works perfectly fine (DA Manager1) -There were one or two people saying that it is not accurate and that it is not working. But we know exactly from our experience that it may not be 100% accurate, there may be some calls that are not recorded correctly but these are some specific scenarios that we found. So it should be, let's say 98% accurate which is enough for us (DA Manager 2) - If the data are not accurate enough because there are issues in something, process, way of working or whatever, not necessarily the issue is real, on the field. We are again challenging the model and the report (DA Manager1)
	Distrust is transferred from algorithms to the DA team	**DA team get blamed if the algorithms are perceived to be faulty** - [DA Employee 1] laughs and says it is like a hot potato. That if anything goes wrong Operation will throw the hot potato to the Data Analytics team and say it was their fault. [DA Employee 2] agrees, that Operations blame Analytics, and this is the result for more processes being automated/using data. (Observation, Fieldnotes) - Most of my colleagues quit working with [DA] team and try to do their own analysis in Excel and using graphs in there and so on. And I think that for part of my colleagues this frustration remained, and they are quite reluctant to go into

Aggregated dimensions	Second order theme	First-order codes and illustrative quotes
		[DA] team and say "OK we need some help"[…] it started I think at the beginning when the [DA team] came in our organization and it was said that "Ok, they can help us, they can provide us with support in doing some things to reduce our manual work. And the people were excited about it, but at the end, when they saw that the things are not going on as they expect it, came this frustration (Operation Employee 2)
		DA team tries to influence operation impact on data, and increase transparency to avoid blame
		- We have right now a way of monitoring the data availability and data quality, so we can raise it before [it becomes an issue with the users of the algorithms]. So, as soon as we observe something is missing and so on, we tried to be proactive and announced our customers "look, it seems that we have some data issue, availability issue, we will delay with this week, analytics will come to you as soon as we get the data back" (DA Employee 7)
		- We started to communicate better and to show that even in Analytics what are the steps that we are taking and what type of reports we are producing […] by doing that with the high recurrence in the beginning, and then with moderate, to show progress, things cooled down (DA Employee 1)
		- So there is kind of a shadowing period when we are trying to look at the model and the result in parallel with the people who also have other ways of looking at data, and we need to see if we reach the same results, and then they also start to understand and to rely on that, because they understand that they model provides the correct data. (DA Manager 2)

Paper 3.

Blinded by the Person? Experimental Evidence from Idea Evaluation

Authors:
Linus Dahlander, Arne Thomas, Martin W. Wallin, Rebecka C. Ångström

Status:
Published in *Strategic Management Journal*

Received: 11 June 2020 | Revised: 1 March 2023 | Accepted: 6 March 2023

DOI: 10.1002/smj.3501

SHORT ARTICLE

STRATEGIC MANAGEMENT JOURNAL WILEY

Blinded by the person? Experimental evidence from idea evaluation

Linus Dahlander[1] | **Arne Thomas**[2] | **Martin W. Wallin**[3,4] | **Rebecka C. Ångström**[5]

[1]ESMT Berlin, Berlin, Germany

[2]University of Amsterdam, Amsterdam Business School, Amsterdam, The Netherlands

[3]Department of Technology Management and Economics, Chalmers University of Technology, Gothenburg, Sweden

[4]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[5]Stockholm School of Economics, House of Innovation, Stockholm, Sweden

Correspondence
Linus Dahlander, ESMT Berlin, Schlossplatz 1, 10178 Berlin, Germany.
Email: linus.dahlander@esmt.org

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: 411843692; Institute for Management of Innovation and Technology (IMIT)

Abstract

Research Summary: Seeking causal evidence on biases in idea evaluation, we conducted a field experiment in a large multinational company with two conditions: (a) blind evaluation, in which managers received no proposer information, and (b) non-blind evaluation, in which they received the proposer's name, unit, and location. To our surprise—and in contrast to the preregistered hypotheses—we found no biases against women and proposers from different units and locations, which blinding could ameliorate. Addressing challenges that remained intractable in the field experiment, we conducted an online experiment, which replicated the null findings. A final vignette study showed that people overestimated the magnitude of the biases. The studies suggest that idea evaluation can be less prone to biases than previously assumed and that evaluators separate ideas from proposers.

Managerial Summary: We wanted to find out if there were biases in the way managers evaluate ideas from their employees. We did a field experiment in a large multinational technology company where we tested two different ways of evaluating ideas: one where managers did not know anything about the person who came up with the idea and one where they did know the person's name, which unit they worked for, and where they were located. The results were surprising. We did not find any bias against women and employees that did not work in the same location and unit as the evaluator. Managers are advised that hiding the identity of idea proposers (from idea evaluators) may not be a silver bullet to improving idea evaluation.

KEYWORDS

bias, field experiment, idea evaluation, innovation, online experiment

1 | INTRODUCTION

The literature on idea evaluation cautions that evaluators can be biased toward certain proposers, meaning that the same idea would receive different evaluation scores depending on who proposed it. Indeed, evaluators often do not base their evaluation solely on the *idea itself* but also on *whose idea* it is (Fuchs, Sting, Schlickel, & Alexy, 2019; Menon & Blount, 2003; Reitzig & Sorenson, 2013). Prior work on idea evaluation has, for example, explained how biases could arise from hierarchy (Keum & See, 2017; Schweisfurth, Schöttl, Raasch, & Zaggl, 2023), sequence (Bian, Greenberg, Li, & Wang, 2021; Criscuolo, Dahlander, Grohsjean, & Salter, 2021), and nepotism (Reitzig & Sorenson, 2013). Knowing who proposed an idea can provide important information (Chaiken, 1980; Pornpitakpan, 2004), yet relying on such source-based heuristics can lead to biases that disadvantage women and people far away from the decision-makers (Banaji & Hardin, 1996; Blair & Banaji, 1996; Stangor, Lynch, Duan, & Glas, 1992). We focus on three potential biases, namely that evaluators provide systematically lower evaluation scores to (a) female idea proposers; and to idea proposers from (b) other units, and (c) other locations.

To empirically assess whether and to what degree these biases are at play in idea evaluation, we conducted a field experiment based on a simple intervention: blinding that withholds information about the idea proposer from evaluators. Prior work has speculated that blinding is a light-touch intervention to remove biases from idea evaluation (Grohsjean, Dahlander, Salter, & Criscuolo, 2022). Blinding might mitigate evaluator biases, ensuring that ideas are evaluated on an equal footing, and has been deployed in diverse settings, such as blind auditions, blind recruitment, and (double-)blind academic peer review. To test blinding in idea evaluation, we conducted a preregistered field experiment inside a large multinational company in the information and communication technology sector. We asked innovation managers to evaluate real business ideas from our partner organization. We expected to identify biases and that blinding would reduce them. To our surprise, blinding the evaluators for the idea proposers' identity had no effect. Acknowledging the limitations of the field experiment and improving generalizability, we replicated the results in an online experiment.

2 | THEORY AND HYPOTHESES

Ideas are the seeds of innovation, but not all seeds bear fruit. It is inherently challenging to assess the potential of new ideas because they are surrounded by market and technological uncertainty. When deciding on new ideas, organizations are likely to make costly errors in the form of false positives (investing in ideas that ultimately fail) and false negatives (missing out on ideas that ultimately become a hit). Ideas' uncertainty can lead to evaluation biases distorting organizational outcomes (Criscuolo et al., 2021; Keum & See, 2017; Reitzig & Sorenson, 2013). Just as idea proposers systematically overestimate the value of their ideas (Fuchs et al., 2019), innovation managers make mistakes in evaluating ideas (Boudreau, Guinan, Lakhani, & Riedl, 2016; Criscuolo, Dahlander, Grohsjean, & Salter, 2017). One challenge for evaluators is to separate ideas from the person who generated them; some idea proposers get the benefit of the doubt, whereas others struggle to get recognized despite having a good idea. For instance, prior work suggests that women and proposers from other units and locations receive lower idea evaluations (see, e.g., Reitzig & Sorenson, 2013).

Given such biases in idea evaluation, we considered a simple intervention to hide the identity of the proposer through blinding. Work on blinding is not new. An influential study on blinding comes from Goldin and Rouse (2000), often quoted to show that the introduction of blind auditions—a blind screen between the jury and an auditioning musician—increased the admission of women to music schools.[1] Most work on blinding comes from studies in academia, where it is a common practice when evaluating papers and grant applications. Early work by Blank (1991, p. 1042) argues that double-blind reviewing in academia "minimizes undesirable referee bias." More recent work by Kolev, Fuentes-Medel, and Murray (2019) found that female authors received lower scores on their grant proposals to the Gates Foundation, even with blinding. Controlling for applicant quality and proposal text suggests biases at a fundamental level, punishing women for producing a different *type* of research rather than research of lesser quality. Evidence from academia thus shows that blinding has the potential to alleviate some biases but not necessarily all types of biases. Inspired by such work on blinding in academia, research on idea evaluation has speculated that blinding could also remove biases in corporate idea evaluation (Grohsjean et al., 2022), yet the evidence to date is scarce.

2.1 | Main effect: Blinding in idea evaluation

Blinding can affect idea evaluation because evaluators may rely on social cues about the person proposing the idea as a heuristic device (see, e.g., Gigerenzer & Gaissmaier, 2011). Heuristics provide mental shortcuts that can save effort by focusing only on the issue's most relevant aspects and ignoring other information. While effective in some regards, heuristics can bias decision-making (Tversky & Kahneman, 1974). Source-based heuristics have been investigated in research on information processing, demonstrating how a source's attributes influence how information is perceived and valued (e.g., Chaiken, 1980; Pornpitakpan, 2004).

The tendency to use information about an idea proposer as a signal for idea quality may be reinforced when evaluators lack information, expertise, or resources to assess an idea's details.

[1]After adding controls for musicians, the study yields mixed results. In some stages, women did worse in blind auditions, which is explained by a potential drop in the quality pool of female candidates after adopting blind auditions. It also shows that the results are more nuanced than often cited (see also Gelman, 2019).

Source-based evaluation heuristics can arise when evaluators have too little information *and* when they have too much. Information overload and time pressure can induce people to rely on simple heuristics (Hansen & Haas, 2001). From the evaluator's perspective, idea evaluation is both information-deprived and information-overloaded—deprived because of the small and standardized information bits that idea descriptions typically hold; overloaded because of the large number of ideas (Piezunka & Dahlander, 2015).

While the effects of blinding are contingent on *what* is blinded (elaborated in the moderation effects below), research suggests that its baseline effect is negative. For instance, acceptance rates are lower and referee reports are less favorable when academic reviewers do not know who the authors are (Blank, 1991; Okike, Hug, Kocher, & Leopold, 2016), and customers evaluated products more favorably when having identity-revealing information on the seller (Forman, Ghose, & Wiesenfeld, 2008). Such findings could be driven by (perceived or actual) selection into blinding by which worse authors or less-trustworthy sellers choose to be blinded. Similarly, the evaluator may perceive identity-revealing information on the idea proposer as a positive signal that the proposer is committed and serious about the idea. Blinding proposer information should thus lead to lower evaluation scores.

Hypothesis (H1). Innovation managers rate ideas lower in blind evaluation.

2.2 | Moderation effects: Who benefits from blinding in idea evaluation?

Studies of academic reviews and hiring decisions suggest that blinding can alter evaluations and potentially overcome biases. We focus on three characteristics that research has found important for idea evaluation: gender, same unit, and shared location (see, e.g., Criscuolo et al., 2017, Reitzig & Sorenson, 2013). Below we elaborate on our expectations of lower scores for women and higher scores for proposers from the same unit and location (as the evaluator), which would disappear if the idea proposer's identity was blinded.[2]

2.3 | Blind to help women in idea evaluation?

Much research documents a tendency to evaluate men and women differently (e.g., Brooks, Huang, Kearney, & Murray, 2014; Heilman, 2001; MacNell, Driscoll, & Hunt, 2015). Gender is a highly visible characteristic that can compensate for unobserved information (Kunda & Spencer, 2003) and a common way to classify other people that occurs almost instantaneously (Brewer & Lui, 1989; Ridgeway, 2006; Stangor et al., 1992). In this process, gender roles and stereotypes are activated, which results in cognitive bias influencing judgment and evaluation (Ridgeway, 2006). Multiple studies have demonstrated how such biases work unfavorably toward women.

These patterns are particularly strong in the technology sector, where they impair women's chances of receiving entrepreneurial funding (e.g., Kanze, Huang, Conley, & Higgins, 2018),

[2]In the pre-analysis plan, we also theorized about potential effects of evaluation order. While recent research has shown that order can affect evaluations (Bian et al., 2021; Criscuolo et al., 2021), blinding is typically not proposed to overcome them. Therefore, we present arguments on evaluation order in the online appendix (Section A1).

progressing into managerial positions (e.g., Tai & Sims, 2005), and receiving equal pay (e.g., Bamberger, Admati-Dvir, & Harel, 1995). In the technology sector, women are confronted with strong male stereotypes (Del Carpio & Guadalupe, 2022), and these gender stereotypes can create role incongruities that work against them. Overloaded with fast-paced information but lacking granular and contextual information, evaluators risk falling back to decision heuristics, thus activating gender stereotypes and providing lower scores to women. In many ways, evaluating ideas is like evaluating entrepreneurial ventures. Research consistently suggests that ventures led by women are perceived as less viable (Lee & Huang, 2018), and that female entrepreneurs are evaluated worse by angel investors (Becker-Blease & Sohl, 2007; Brooks et al., 2014), venture capitalists (Greene, Brush, Hart, & Saparito, 2001; Nelson & Levesque, 2007), and CFOs (Graham & Harvey, 2001). These inequalities arise at least partially from role incongruity between female stereotypes and the images of successful entrepreneurs, although they can be mitigated by other factors such as framing (Lee & Huang, 2018). The same "lack of fit" may handicap women in idea evaluation, where the idea proposer is often expected to develop the idea further as an intrapreneur. We thus hypothesize that women are at a disadvantage compared to men and that blinding would remove this disadvantage.

> **Hypothesis (H2).** Innovation managers rate ideas that are proposed by female employees higher in blind evaluation.

2.4 | Blind to help people outside the unit in idea evaluation?

Scholars have repeatedly demonstrated that people are positively biased toward members of their group (see, e.g., Mullen, Brown, & Smith, 1992). One mechanism underlying in-group bias is that prior interaction increases comfort and reduces objectivity (Lawler, 1992; Zajonc, 1968). The preference for ideas from the same unit can also arise through categorization, as randomly assigning subjects to the same groups induces in-group preferences even in the absence of direct social interaction (Tajfel, Billig, Bundy, & Flament, 1971). In-group bias can arise for strategic reasons, where the evaluator looks to further his/her organizational unit. Pushing ideas from the own unit forward within the organization can bring additional resources and increase prestige, and successful ideas can help achieve the unit's strategic or business targets, which benefit the evaluator, especially if holding a managerial role. In-group bias can also arise because evaluators subconsciously prefer ideas from their unit. Evaluators likely perceive ideas from the same unit as more interesting and understandable because they have a shared understanding of key challenges, technologies, and market needs. Strategic considerations and subconscious processes can increase the scores evaluators give to ideas proposed by one of their own. In line with these arguments, Reitzig and Sorenson (2013) demonstrated that middle managers are biased in favor of ideas from their division. These findings align with the not-invented-here syndrome, in which groups believe they have a monopoly on knowledge, reject outside ideas, and promote their unit (Katz & Allen, 1982). While we expect in-group biases in idea evaluation, blinding counteracts them and thus reduces evaluation scores for ideas originating in the same unit as the evaluator.

> **Hypothesis (H3).** Innovation managers rate ideas that are proposed by employees from the same organizational unit lower in blind evaluation.

2.5 | Blind to help people outside the location in idea evaluation?

Being in the same unit does not always imply sharing a location, which indicates that location and unit need to be analytically separated. However, similar arguments apply. A shared location typically implies a greater mutual understanding of cultural aspects and speaking the same language (metaphorically or literally). This makes ideas from colocated proposers and evaluators more relatable and accessible, reducing the cognitive burden when evaluating ideas. Moreover, identification and a sense of togetherness among colocated employees may trigger favoritism, like in-group bias. Feeling a closer emotional association with colleagues from the same location, evaluators may pay special attention to their ideas or give them the benefit of the doubt. Research on idea evaluation has found such tendencies to favor colleagues from the same location (Reitzig & Sorenson, 2013). Studies in international business have illustrated this at the firm-level, where a home country bias in R&D activities exists (Belderbos, Leten, & Suzuki, 2013). Blinding may thwart this inclination, resulting in lower evaluation scores for ideas from the same location.

> **Hypothesis (H4).** Innovation managers rate ideas that are proposed by employees located in the same country lower in blind evaluation.

3 | A FIELD EXPERIMENT IN A MULTINATIONAL COMPANY

Our field experiment tests the hypotheses in a real-world setting where managers have a stake in their decisions. We asked innovation managers at our partner company—a leading multinational company in the information and communication technology sector—to evaluate employee ideas proposed through the company's idea management system. We experimentally varied whether information on the idea proposer was blinded or not. The experiment was preregistered at the American Economic Association RCT Registry under the ID AEARCTR-0005439.[3]

3.1 | Participants

We recruited participants from the formal network of innovation managers operated by our industrial partner's internal accelerator. Like regular venture capitalists, the accelerator looks to develop employee ideas into new businesses and offers intrapreneurs funding and access to company personnel, expertise, and partnerships. The accelerator's process has several stages. Our field experiment is situated at the very beginning of that process—where innovation managers review and evaluate employees' initial ideas. The network of innovation managers consists of volunteers and all employees can apply, irrespective of position, unit, or location. Once accepted into the network, employees go through a short, formalized training on (a) user-generated, design-driven innovation, (b) the company's idea management system, and (c) business

[3]Note that we preregistered our study after the experiment started, but before we retrieved or inspected any data. The preregistration can therefore be classified as a registration prior to any human observation of the data (as defined in the OSF preregistration template, available at https://osf.io/prereg).

coaching. One task of the newly trained innovation manager is to support idea proposers in developing and improving their ideas. Most importantly for our study, the innovation managers regularly evaluate and give feedback on early ideas and recommend mature ideas for the company's internal funding process. When making a recommendation, they use the same criteria we employ in the experiment (see Section 3.5). Sixty innovation managers signed up for our experiment, thirty-eight completed it (63.3%), and eight evaluators (13.3%) started the idea evaluation but did not finish it.

Neither the innovation managers evaluating the ideas nor the employees proposing them were aware that their evaluations or ideas were part of an experiment. Instead, we communicated—in line with the messages of the internal accelerator—that the evaluations were an additional input to the company's effort to unlock the intrapreneurial spirit and improve idea evaluation. Even our contact persons were not aware of our exact research interest. We took great care not to reveal our research question or experimental manipulation to avoid experimenter demand effects (Rosenthal, 1966).

3.2 | Evaluation task

We used the survey tool Qualtrics to design an online evaluation interface, customizing its flow and visual appearance. We mirrored our industrial partner's corporate design to maximize the integrity and credibility of the online idea evaluation as an important organizational task. After a brief welcome screen, each idea was presented on an individual evaluation screen containing (in this order): a short request to evaluate the idea, information on the idea proposer depending on the treatment (see Section 3.3), the idea title, the idea description, five questions to rate the ideas (see Section 3.5), and a text field for open comments. We provide a stylized illustration of the idea evaluation screen in Figure 1 and the survey flow in Figure A1.[4]

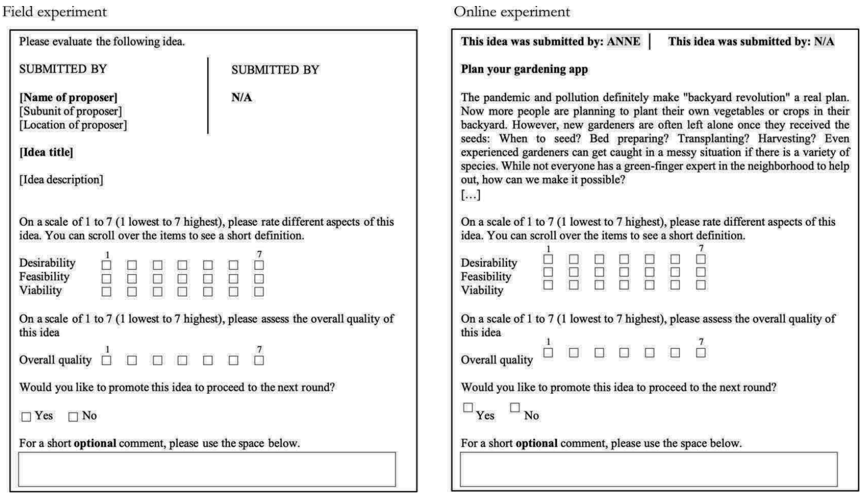
Each innovation manager was asked to evaluate 48 ideas. The ideas came from a larger pool of ideas through the company's idea management system, an online platform where employees can submit ideas and interact with others to refine them. It is an important tool in our partner's innovation process. Ideas are evaluated regularly, and there is a budget specifically for their development. Evaluating these ideas is thus the first step toward possible larger, impactful investments down the line.

For our experimental manipulation (blinding) to work credibly, we needed early stage ideas unknown to the participants. Therefore, we considered only the 412 ideas proposed in the 6 months before our study. We left titles and descriptions unchanged.[5] In terms of content, most of the ideas were categorized under four headlines: autonomous vehicles (124 ideas), design thinking (87 ideas), logistics (86 ideas), and smart manufacturing (64 ideas). On average, the ideas had 120.24 words and received 3.21 comments. This shows the ideas had not received much attention prior to our experiment. We cannot share idea details or concrete examples because they are proprietary.

We received 1,942 idea evaluations; 1,824 (38 evaluators × 48 ideas) from innovation managers who finished the idea evaluation, and 118 from those who did not finish. We excluded

[4]Sections, tables, and figures with an "A" (e.g., Section A3, Table A1, Figure A7) are in the online appendix.

[5]In total, we retrieved 570 ideas that had been submitted between February 6, 2019, and October 7, 2019. Besides restricting the time frame (from April 8, 2019, to October 7, 2019), we took additional steps in selecting ideas. These steps are described in Section A2.


Field experiment

Please evaluate the following idea.

SUBMITTED BY

[Name of proposer]
[Subunit of proposer]
[Location of proposer]

SUBMITTED BY

N/A

[Idea title]

[Idea description]

On a scale of 1 to 7 (1 lowest to 7 highest), please rate different aspects of this idea. You can scroll over the items to see a short definition.

Desirability 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ 7
Feasibility ☐ ☐ ☐ ☐ ☐ ☐ ☐
Viability ☐ ☐ ☐ ☐ ☐ ☐ ☐

On a scale of 1 to 7 (1 lowest to 7 highest), please assess the overall quality of this idea

Overall quality 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ 7

Would you like to promote this idea to proceed to the next round?

☐ Yes ☐ No

For a short **optional** comment, please use the space below.

Online experiment

This idea was submitted by: ANNE | **This idea was submitted by: N/A**

Plan your gardening app

The pandemic and pollution definitely make "backyard revolution" a real plan. Now more people are planning to plant their own vegetables or crops in their backyard. However, new gardeners are often left alone once they received the seeds: When to seed? Bed preparing? Transplanting? Harvesting? Even experienced gardeners can get caught in a messy situation if there is a variety of species. While not everyone has a green-finger expert in the neighborhood to help out, how can we make it possible?
[…]

On a scale of 1 to 7 (1 lowest to 7 highest), please rate different aspects of this idea. You can scroll over the items to see a short definition.

Desirability 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ 7
Feasibility ☐ ☐ ☐ ☐ ☐ ☐ ☐
Viability ☐ ☐ ☐ ☐ ☐ ☐ ☐

On a scale of 1 to 7 (1 lowest to 7 highest), please assess the overall quality of this idea

Overall quality 1 ☐ ☐ ☐ ☐ ☐ ☐ ☐ 7

Would you like to promote this idea to proceed to the next round?

☐ Yes ☐ No

For a short **optional** comment, please use the space below.

FIGURE 1 Stylized idea evaluation screen of field experiment and online experiment (non-blind condition left and blind condition right).

seven evaluations that were completed in less than 8 s. Our main sample contained 1,837 idea evaluations because in 98 cases the innovation managers did not rate the ideas on our main dependent variable.

3.3 | Treatment conditions

We used two conditions: *blind evaluation*, in which the innovation manager received no information about the idea proposer ("Submitted by: N/A"), and *non-blind evaluation*, in which the innovation manager received information about the idea proposer (name, unit, geographical location). We used a *within-subject design* in which each innovation manager evaluated ideas under both conditions.

3.4 | Randomization

Each innovation manager evaluated 48 ideas, which we randomly picked from the idea pool. To ensure each innovation manager evaluated ideas from employees with diverse backgrounds, we relied on stratified random assignment. Each idea was assigned to one of 20 strata based on the proposer's gender (2 strata) and unit (10 strata). We then randomly picked ideas from each stratum using a built-in function in Qualtrics. The number of ideas picked from each stratum was roughly proportional to the stratum size, although we oversampled small strata. After ideas were randomly picked, we randomized the evaluation order and blinded the evaluators at random.

3.5 | Variables

3.5.1 | Dependent variables

As specified in the pre-analysis plan, we used *Overall score* as our main dependent variable. For each idea, evaluators were asked: "On a scale of 1 to 7 (1 lowest to 7 highest), please assess the overall quality of the idea." In robustness checks, we used alternative-dependent variables regularly used at our partner firm (see Section A3).

3.5.2 | Treatment variable

We used an indicator variable *Blind* that switched to 1 if an evaluator evaluated an idea in the blind condition.

3.5.3 | Moderator variables

To test (H2), *Female proposer* switched to 1 if an idea was proposed by a woman. We coded gender based on name matching. If name matching did not yield a clear match, we consulted internal documents from our partnering firm to resolve ambiguity.

To test (H3), *Same unit* switched to 1 if an idea proposer was part of the same division, function, or geographical market as the innovation manager evaluating the idea. At our industrial partner, each employee belongs to one and only one division, function, or geographical market. We determined the employee's position based on an internal code, which we retrieved from the company's intranet.

To test (H4), *Same location* switched to 1 if an idea proposer was in the same country as the evaluator. We coded the location from the employees' addresses in the internal records.

4 | RESULTS

Table A1 provides an overview of the variables and their correlations. It shows that 50.2% of the ideas were rated in a blind evaluation and that 16.7, 12.9, and 9.5% came from a female proposer, the same unit, and the same location, respectively. On average, the ideas received an *Overall score* of 3.32. There was only modest consensus among the evaluators. In 20.75% of the cases, they agreed on the rating and the intraclass correlation coefficient for the ideas was 0.1496 (one-way random effects model).

As a first step, we conducted mean comparisons between the treatment conditions. Table 1 reveals that the difference was small (0.0636), and *t* tests failed to reject that blind and non-blind evaluation produce the same mean outcomes. For innovation outcomes, it is also important to consider extreme outcomes in the tails of the distribution. Figure A2 shows that the distributions of *Overall score* exhibit no clear differences between the blind and non-blind condition, either in the middle or in the tails. Overall, these first results provide no support for (H1).

To test our hypotheses more conclusively, we ran a series of ordinary least squares estimations that linked the *Overall score* that idea i received from innovation manager j to a treatment

TABLE 1 Mean comparison of treatment conditions in field and online experiment

DV: Overall score	Field experiment	Online experiment—Within design	Online experiment—Between design
Non-blind condition	3.353	4.666	4.575
Blind condition	3.289	4.608	4.587
Difference	0.0636	0.0580	−0.0123
t-statistic	0.90	1.62	−0.34
p-value	0.37	0.10	0.74
N	1,837	7,332	7,331

indicator *Blind* and fixed effects for the idea, the evaluator, and the display order. For (H2) to (H4), we also included variables indicating ideas from a *Female proposer*, the *Same unit*, and the *Same location* and a series of interaction terms with the treatment indicator. Because we could not separately blind one characteristic of an idea proposer (e.g., gender), we estimated their effects in one model[6]:

$$\begin{aligned} \textit{Evaluation score}_{ij} = {} & \beta_0 + \beta_1 \textit{Blind}_{ij} + \beta_2 \textit{Female proposer}_i + \beta_3 \textit{Same subunit}_{ij} \\ & + \beta_4 \textit{Same location}_{ij} + \beta_5 \textit{Blind}_{ij} * \textit{Female proposer}_i + \beta_6 \textit{Blind}_{ij} \\ & * \textit{Same subunit}_{ij} + \beta_7 \textit{Blind}_{ij} * \textit{Same location}_{ij} + \beta_8 \textit{Idea FE}_i \\ & + \beta_9 \textit{Evaluator FE}_j + \beta_{10} \textit{Order FE}_{ij} + \varepsilon_{ij}. \end{aligned}$$

Table 2 reports the regression results. Despite strong theoretical priors, we found no support for any of the hypotheses. On average, innovation managers rated the ideas only 0.0989 points lower (95% CI [−0.2241, 0.0264]) in the blind evaluation, providing no support for (H1). Regarding (H2), ideas from female proposers were rated only 0.109 points higher (95% CI [−0.1877, 0.4063]) in the blind evaluation. Regarding (H3), overall scores for ideas from the same unit exhibited practically no difference between the treatment conditions (point estimate of 0.0258; 95% CI [−0.4166, 0.4682]). Regarding (H4), ideas from the same location were rated even higher in the blind condition (by 0.145 points; 95% CI [−0.2493, 0.5387]). Overall, we found no support for our hypotheses in any of the three different analytical approaches: (a) mean comparisons and *t* tests, (b) visual inspection of distribution graphs, and (c) regression analyses.

4.1 | Post hoc analyses: Exploring and replicating the null finding

In line with *SMJ*'s guidelines, we assessed the robustness of our null finding in post hoc analyses using several alternative operationalizations, samples, and estimation techniques (see Section A4). The results hold when accounting for (a) alternative dependent variables, (b) differences in evaluation effort, (c) sample size and power, (d) idea quality, (e) distributional effects, (f) demand effects, (g) alternative time thresholds, and (h) experimental hurdles.

[6]We used the outlined model to test (H2) to (H4). To test (H1), we restricted the model to include only *Blind* and the fixed effects.

TABLE 2 Regression analyses for field and online experiment

DV: Overall score	Field experiment (1)	Field experiment (2)	Online experiment—Within design (3)	Online experiment—Within design (4)	Online experiment—Between design (5)	Online experiment—Between design (6)
H1: Blind	−0.0989 (.119)	−0.133 (.0770)	−0.0629 (.0497)		0.0124 (.717)	
Female proposer		−0.401 (.583)				
Same subunit		0.170 (.280)				
Same location		−0.235 (.211)				
H2: Blind × female proposer		0.109 (.462)				
H3: Blind × same subunit		0.0258 (.907)				
H4: Blind × same location		0.145 (.463)				
H2: Female name				0.0696 (.0877)		−0.00273 (.948)
H2: Male name				0.0562 (.129)		−0.0221 (.603)
Constant	2.696 (.000)	2.703 (.000)	4.947 (.000)	4.884 (.000)	4.860 (.000)	4.873 (.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	No	No
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,837	1,837	7,314	7,314	7,304	7,304
R^2	.4049	.4074	.1176	.1176	.1066	.1066

Note: Experiment and experimental design are given in the column headers. *p*-Values in parentheses are based on *SE*s clustered at the evaluator level. In the online experiment, we randomly assigned female and male names to the ideas. Therefore, we did not need to specify an interaction. Instead, the main effects of *Female name* and *Male name* give the difference between assigning female and male names and the blind condition, our baseline.

We conducted an additional online experiment on Prolific to address intractable challenges from the field experiment, detailed in Section A5. We replicated the field experiment as closely as possible, while also addressing its limitations. First, we increased the sample size and statistical power by recruiting 1,543 participants. Second, we reduced the number of ideas per evaluator to 10 and incentivized the participants with above-average compensation (yielding 14,663 evaluations). Third, we ensured high and constant idea quality by using successful ideas from crowdfunding and standardizing them in terms of length and presentation. Fourth, we replicated the within design of the field experiment and complemented it with a between design (in which evaluators evaluate only blinded or only non-blinded ideas). Fifth, we focused on the proposer's gender, the most meaningful aspect of the proposer's identity in the online setting and the strongest effect in the field experiment. The evaluation setup, experimental conditions, and randomization closely resembled the field experiment.

We took the same steps to analyze the data as in the field experiment: mean comparisons, distribution graphs, and regression analyses. Table 1 shows that the mean differences were small (0.0580 and −0.0123) and t tests failed to reject that blind and non-blind evaluations produce the same mean outcomes.[7] Figure A3 shows no notable differences between blind and non-blind evaluations and between the two experimental designs. Table 2 shows that blind evaluations were slightly lower in the within design (−0.0629; 95% CI [−0.1257, −0.0001]), but there was no difference in the between design (0.0124; 95% CI [−0.0549, 0.0798]).[8] In neither design did we find differences in the evaluation scores of ideas proposed by women or men, compared to those in the blind condition. Overall, these results thus confirm the null finding from the field experiment. A final vignette study, detailed in Section A6, showed that independent participants overestimated gender bias in idea evaluation.

5 | DISCUSSION

We conducted a field experiment at one of the world's leading technology firms, where we randomly assigned innovation managers to evaluate ideas in a blind and non-blind condition. Prior research had identified biases that could distort the evaluation of ideas (e.g., Boudreau et al., 2016; Criscuolo et al., 2017; Reitzig & Sorenson, 2013) and suggested blinding as a light-touched intervention to remove such biases (Grohsjean et al., 2022). We build on work that has identified gender, organizational structure, and location as sources of bias, but found no differences between blind and non-blind evaluations. This null finding was thoroughly assessed in post hoc analyses, replicated in an online experiment, and contributes to building cumulative knowledge in strategic management (Bettis, Ethiraj, Gambardella, Helfat, & Mitchell, 2016).

A fundamental question for strategy researchers is how to allocate resources to innovation projects (see, e.g., Criscuolo et al., 2017; Klingebiel & Rammer, 2014; Reitzig & Sorenson, 2013). These decisions have large consequences for companies, prompting scholarly work on how they are made and how to keep biases out of the process. We provide experimental evidence from one of the world's leading technology firms. Indeed, the lack of causal evidence from within-company settings limits scholarly understanding of innovation. Most experimental research on biases and the evaluation of intellectual work has been conducted in *non*-corporate settings, particularly in academic settings (see, e.g., Blank, 1991) that may limit the application of prior findings. One reason for the lack of corporate field experiments is that companies are generally reluctant to grant access to internal evaluation processes and provide data on the organization's choice set of ideas. Our industrial research partner opened their doors and provided us with an unfiltered set of ideas they had not previously evaluated, preventing selective sampling and success bias.

[7]Note that *Overall score* was slightly higher in non-blind evaluation for the within design, with a marginally statistically significant difference ($p = .10$). The mean comparisons also support the appropriateness of the within design used in the field experiment, as the average evaluation scores are very similar for the two experimental designs.
[8]Although statistically significant, the main effect of blinding in the within design is small and hardly economically meaningful. Its statistical significance is largely the result of a larger sample.

Considering our experimental findings and setup, we elaborate on four plausible reasons for our null finding: (a) organizational culture, (b) selection into the experiment, (c) separation of idea from person, and (d) shifting standards from evaluation to selection.

First, the null finding might be due to organizational culture. When we shared our results with our industrial partner, some explained them with an engineering culture in which "ideas matter more than people." Our partner firm is a prestigious employer scouting globally for the best engineering talent and using standardized hiring and selection policies, which ensures fit with organizational culture and homogenous evaluators. However, the online experiment replicated the field experiment but removed the engineering culture and increased variation in evaluator quality. The null finding persisted, suggesting that the organizational culture of our industrial partner cannot explain it.

Second, selection into the experiment could lead to the null finding. Studies of hiring discrimination, for instance, have found that blinding can make it harder for members of outgroups to be hired. Blinding may prevent "positive" discrimination, in which recruiters look, for example, to increase the number of women but can no longer be more generous toward them (see, e.g., Behaghel, Crépon, & Le Barbanchon, 2015). The risk for our study is that people positively inclined to give women and members of the outgroup higher evaluations would select to be part of our experiment. However, the firm identified the group of innovation managers participating in our experiment (limiting self-selection into the experiment). There is also no opportunity to select into the online experiment based on being more lenient toward disadvantaged groups. This makes us conclude that selection into the experiment cannot explain the null finding.

Third, blinding may be more effective when the person's identity is more tied to information deemed critical for assessing ideas. Our null finding contrasts CV experiments in which women with identical CVs often receive lower evaluations than men (e.g., Petit, 2007), and evidence from entrepreneurship, where woman-led ventures are perceived as less viable (Lee & Huang, 2018) and female-backed female entrepreneurs receive lower evaluations than men (Snellman & Solal, 2023). In all these cases, the evaluation process is at least as much about the person as it is about the idea, which may differ in our setting. Our null finding may arise because the idea takes precedence over the person; the proposer's identity does not evoke information deemed critical to idea evaluation. Even when we made the proposer's gender more salient in the online experiment, we found no gender differences. Based on these findings, there are good reasons to believe that blinding is not guaranteed to improve idea evaluation. It may be ineffective if the idea proposer's identity does not unlock strong schemas that blinding could curtail.

Fourth, people may apply different standards when evaluating than when selecting ideas, and biases only manifest themselves in the selection. The "shifting standards model" in social psychology (Biernat & Manis, 1994) suggests such a difference between evaluation and selection. For instance, when evaluating job candidates, a female candidate may be seen as "good for being a woman." However, in selection decisions where candidates (or ideas) are pitted against each other as there are limited resources, there are usually greater biases (Joshi, Son, & Roh, 2015; Koch, D'Mello, & Sackett, 2015). We study the evaluation of early stage ideas, which still have a long way to go to eventual selection. The evaluators do not make the final selection and have few budget constraints, which may reduce biases. This could explain why our results differ from previous work focused on selection (e.g., Reitzig & Sorenson, 2013). This opens the question of how to design evaluation and selection, as the ultimate selection requires favorable evaluations along the way.

We invite future research to explain why our surprising null finding occurs. We acknowledge that the null finding is more robust for gender than that for organizational unit and location because we have replicated it in the online experiment. Future research on blinding is warranted to help explain our null finding. Our theorizing distinguished between strategic favoritism and subconscious preferences. Blinding can only be effective against biases caused by the withheld information. We blinded information on the idea proposer but left idea descriptions unchanged. To the degree to which biases are prompted by, for example, more interesting, exciting, and accessible idea descriptions, blinding is ineffective. Similarly, blinding could be ineffective because the evaluator has learned about an idea before, a common complication for academic peer review, where reviewers may have seen the paper presented at a conference. The same can happen with ideas from the same unit or location. Future research could further investigate such boundaries to effective blinding and the relative importance of strategic vs. subconscious biases. Given that blinding did not improve idea evaluation in our experiment, future research should investigate the costs of blinding. While blinding ideas is technically straightforward and relatively costless to implement, missing out on blocked information has opportunity costs. It reduces the potential to connect employees with similar interests and learn from what other people are working on. After all, evaluation is not an end goal for companies but one of many steps from an idea to a successful product. Blinding is potentially most helpful on a smaller scale to determine whether, how, and where biases exist before scaling any idea evaluation initiative within or across organizations.

Finally, the design and implementation of blinding may be less straightforward and more context dependent than most studies assume. Our field experiment and the follow-up online experiment are both "online," in the sense of displaying information as text-on-screen. This is a common practice to collect and assess ideas (Bayus, 2013; Beretta, 2019; Blohm, Riedl, Füller, & Leimeister, 2016; Poetz & Schreier, 2012), and future research will need to investigate potential differences between traditional idea evaluation where evaluation panels and idea proposers can meet up at the corporate headquarters and online idea evaluation—especially in terms of the schemas they unlock, and the potential to blind the information that prompts their unraveling.

ACKNOWLEDGMENTS

All authors contributed equally and are listed alphabetically. The authors are grateful for comments from Oliver Alexy, Gianluca Carnabuci, Florian Englmaier, Christoph Fuchs, Thorsten Grohsjean, Marcus Holgersson, Reddi Kotha, Tobias Kretschmer, Friederike Lenel, Markus Reitzig, David Ronayne, Ammon Salter, Henry Sauermann, Isabelle Solal, Martin Schweinsberg, Dashun Wang, and seminar participants at the Duke University field experimental workshop, Bath University, Ludwig Maximilian University of Munich, Singapore Management University, Stockholm School of Economics, TU Munich, Chalmers University of Technology, and the Imperial community. Hai Anh Vu provided outstanding research assistance. The project is funded by Deutsche Forschungsgemeinschaft (DFG) under project number 411843692 and the Institute for Management of Innovation and Technology (IMIT). The authors claim that all the errors are theirs alone. Open Access funding enabled and organized by Projekt DEAL.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data from the two online experiments is available at [https://osf.io/dkbpx/].

DATA AVAILABILITY STATEMENT

The field experiment is under NDA and we cannot share the data. The data from the online experiments can be fully shared (including the code). We would be very happy to share this data if anyone is interested!

REFERENCES

Bamberger, P., Admati-Dvir, M., & Harel, G. (1995). Gender-based wage and promotion discrimination in Israeli high-technology firms: Do unions make a difference? *Academy of Management Journal*, *38*(6), 1744–1761.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*(3), 136–141.

Bayus, B. L. (2013). Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science*, *59*(1), 226–244.

Becker-Blease, J. R., & Sohl, J. E. (2007). Do women-owned businesses have equal access to angel capital? *Journal of Business Venturing*, *22*(4), 503–521.

Behaghel, L., Crépon, B., & Le Barbanchon, T. (2015). Unintended effects of anonymous resumes. *American Economic Journal: Applied Economics*, *7*(3), 1–27.

Belderbos, R., Leten, B., & Suzuki, S. (2013). How global is R&D? Firm-level determinants of home-country bias in R&D. *Journal of International Business Studies*, *44*(8), 765–786.

Beretta, M. (2019). Idea selection in web-enabled ideation systems. *Journal of Product Innovation Management*, *36*(1), 5–23.

Bettis, R. A., Ethiraj, S., Gambardella, A., Helfat, C., & Mitchell, W. (2016). Creating repeatable cumulative knowledge in strategic management: A call for a broad and deep conversation among authors, referees, and editors. *Strategic Management Journal*, *37*(2), 257–261.

Bian, J., Greenberg, J., Li, J., & Wang, Y. (2021). Good to go first? Position effects in expert evaluation of early-stage ventures. *Management Science*, *68*(1), 300–315.

Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*(1), 5–20.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology*, *70*(6), 1142–1163.

Blank, R. M. (1991). The effects of double-blind versus single-blind reviewing: Experimental evidence from the American Economic Review. *American Economic Review*, *81*(5), 1041–1067.

Blohm, I., Riedl, C., Füller, J., & Leimeister, J. M. (2016). Rate or trade? Identifying winning ideas in open idea sourcing. *Information Systems Research*, *27*(1), 27–48.

Boudreau, K. J., Guinan, E. C., Lakhani, K. R., & Riedl, C. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science*, *62*(10), 2765–2783.

Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition*, *7*(3), 262–274.

Brooks, A. W., Huang, L., Kearney, S. W., & Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences*, *111*(12), 4427–4431.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, *60*(2), 433–460.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2021). The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organization Science*, *32*(4), 987–1008.

Del Carpio, L., & Guadalupe, M. (2022). More women in tech? Evidence from a field experiment addressing social identity. *Management Science*, *68*(5), 3196–3218.

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313.

Fuchs, C., Sting, F. J., Schlickel, M., & Alexy, O. (2019). The ideator's bias: How identity-induced self-efficacy drives overestimation in employee-driven process innovation. *Academy of Management Journal*, *62*(5), 1498–1522.

Gelman, A. (2019). Did blind orchestra auditions really benefit women? Retrieved from https://statmodeling.stat.columbia.edu/2019/05/11/did-blind-orchestra-auditions-really-benefit-women/

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*(1), 451–482.

Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, *90*(4), 715–741.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics*, *60*(2–3), 187–243.

Greene, P. G., Brush, C. G., Hart, M. M., & Saparito, P. (2001). Patterns of venture capital funding: Is gender a factor? *Venture Capital*, *3*(1), 63–83.

Grohsjean, T., Dahlander, L., Salter, A. J., & Criscuolo, P. (2022). Better ways to green-light new projects. *MIT Sloan Management Review*, *63*(2), 33–38.

Hansen, M. T., & Haas, M. R. (2001). Competing for attention in knowledge markets: Electronic document dissemination in a management consulting company. *Administrative Science Quarterly*, *46*(1), 1–28.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674.

Joshi, A., Son, J., & Roh, H. (2015). When can women close the gap? A meta-analytic test of sex differences in performance and rewards. *Academy of Management Journal*, *58*(5), 1516–1545.

Kanze, D., Huang, L., Conley, M. A., & Higgins, E. T. (2018). We ask men to win and women not to lose: Closing the gender gap in startup funding. *Academy of Management Journal*, *61*(2), 586–614.

Katz, R., & Allen, T. J. (1982). Investigating the Not Invented Here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R&D project groups. *R&D Management*, *12*(1), 7–20.

Keum, D. D., & See, K. E. (2017). The influence of hierarchy on idea generation and selection in the innovation process. *Organization Science*, *28*(4), 653–669.

Klingebiel, R., & Rammer, C. (2014). Resource allocation strategy for innovation portfolio management. *Strategic Management Journal*, *35*(2), 246–268.

Koch, A. J., D'Mello, S. D., & Sackett, P. R. (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, *100*(1), 128–161.

Kolev, J., Fuentes-Medel, Y., & Murray, F. (2019). *Is blinded review enough? How gendered outcomes arise even under anonymous evaluation* (NBER Working Paper No. 25759). National Bureau of Economic Research, Cambridge, MA.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*(4), 522–544.

Lawler, E. J. (1992). Affective attachments to nested groups: A choice-process theory. *American Sociological Review*, *57*(3), 327–339.

Lee, M., & Huang, L. (2018). Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science*, *29*(1), 1–16.

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291–303.

Menon, T., & Blount, S. (2003). The messenger bias: A relational model of knowledge valuation. *Research in Organizational Behavior*, *25*(3), 137–186.

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103–122.

Nelson, T., & Levesque, L. L. (2007). The status of women in corporate governance in high–growth, high–potential firms. *Entrepreneurship Theory and Practice*, *31*(2), 209–232.

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *Journal of the American Medical Association*, *316*(12), 1315–1316.

Petit, P. (2007). The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector. *Labour Economics*, *14*(3), 371–391.

Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, *58*(3), 856–880.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, *29*(2), 245–256.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281.

Reitzig, M., & Sorenson, O. (2013). Biases in the selection stage of bottom-up strategy formulation. *Strategic Management Journal*, *34*(7), 782–799.

Ridgeway, C. L. (2006). Gender as an organizing force in social relations: Implications for the future of inequality. In F. D. Blau, M. B. Brinton, & D. B. Grusky (Eds.), *The declining significance of gender* (pp. 265–287). New York, NY: Russell Sage.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. New York, NY: Appleton-Century-Crofts.

Schweisfurth, T. G., Schöttl, C. P., Raasch, C., & Zaggl, M. A. (2023). Distributed decision-making in the shadow of hierarchy: How hierarchical similarity biases idea evaluation. *Strategic Management Journal*, 1–28. https://doi.org/10.1002/smj.3497

Snellman, K., & Solal, I. (2023). Does investor gender matter for the success of female entrepreneurs? Gender homophily and the stigma of incompetence in entrepreneurial finance. *Organization Science*, *34*(2), 680–699.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207–218.

Tai, A. J. R., & Sims, R. L. (2005). The perception of the glass ceiling in high technology companies. *Journal of Leadership & Organizational Studies*, *12*(1), 16–23.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*(2, Pt. 2), 1–27.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dahlander, L., Thomas, A., Wallin, M. W., & Ångström, R. C. (2023). Blinded by the person? Experimental evidence from idea evaluation. *Strategic Management Journal*, 1–17. https://doi.org/10.1002/smj.3501

Online appendix.
Blinded by the Person? Experimental Evidence from Idea Evaluation

ONLINE APPENDIX
Blinded by the Person?
Experimental Evidence from Idea Evaluation

A1. Theory: Order effects

Building on the attention-based view, scholars have likened the evaluation of ideas to a tiring muscle—while each idea is scrutinized in the beginning, cognitive fatigue eventually makes it difficult to uphold the same standard (Baumeister, Bratslansky, Muraven, and Tice, 1998). Such order effects are also evident in jury evaluations in sports and entertainment, where late performers regularly receive higher scores (de Bruin, 2005). A possible explanation is that judges form opinions about late performers based on earlier performers, giving more (positive) weight to unique features and making it easier for late performers to stand out (Houston, Sherman, and Baker, 1989; de Bruin and Keren, 2003). In sum, we expect negative order effects of blinding because source-based heuristics become more important when cognitive fatigue takes over, and removing differentiating proposer information hurts ideas more later in the evaluation process.

Hypothesis (A1). *The negative effect of blind evaluation grows stronger for ideas that are evaluated later in the process.*

A2. Field experiment: Idea selection

Employees submitted ideas through our industrial partner's dedicated idea management platform. In total, we retrieved 570 ideas that were submitted between February 6, 2019 and October 7, 2019 to the idea management system. We eliminated 15 duplicate ideas and 5 ideas with blank idea descriptions.

For our experimental manipulation (blinding) to work credibly, we needed early-stage ideas that had not passed through our partner firm and ideally had not be seen by the innovation managers before. To arrive at an appropriate set of ideas, we took the following steps:

1. We restricted the set of ideas to ideas from the past six months (from April, 8 2019 to October 7, 2019) (excluding 39 ideas).
2. We excluded ideas that have already progressed in the internal implementation process, e.g. by having a coach assigned to develop the idea further (5 ideas).
3. We excluded ideas that contained links or other additional material, e.g. a text document or presentation slides (64 ideas).
4. We excluded ideas with missing information on the proposer's gender, organizational unit or location (12 ideas).
5. We excluded ideas that were proposed by one of the innovation managers who evaluated the ideas (18 ideas).

A3. Field experiment: Alternative dependent variables

As specified in the pre-analysis plan and reported in the main text, we used *Overall score* as our main outcome. However, we also used four alternative dependent variables. We asked the evaluators to rate the ideas along three dimensions regularly used at our partner firm: *Desirability* (defined as "The degree to which a solution to a problem addresses someone's needs."), *Feasibility* (defined as "The degree to which a solution is possible and suitable for [partner firm] to implement."), and *Viability* (defined as "The degree to which an idea makes business sense for [partner firm]."). Just as *Overall score*, our main dependent variable, we asked the evaluators to rate ideas along the dimensions on a 7-point scale. Finally, we asked the evaluators whether an idea should be promoted to the next round— *yes* or *no*. The answer was recorded in the dummy variable *Next round*.

A4. Field experiment: Post-hoc analyses

Possible explanations for the lack of statistical support for our hypotheses are problems in the statistical analysis or experimental setup leading to a false null finding. To assess the robustness of our null finding, we addressed several potential concerns regarding our experiment and analysis in post-hoc analyses.[1]

Alternative dependent variables. We considered alternative dependent variables to test our hypotheses. Besides *Next round* as an indicator for overall idea quality, we also used *Desirability*, *Feasibility*, and *Viability* to capture specific aspects of the idea. Table A2 shows that the mean differences between the treatment conditions were small for these alternate variables (the highest difference being 0.0995) and t-tests failed to reject that blind and non-blind evaluations produce the same mean outcomes. The distribution graphs, shown in Figures A4–A7, also exhibit no differences. Finally, Models A1–A4 in Table A3 illustrate that the null finding prevails, although the general effect of blind evaluation is slightly higher on the alternate variables, ranging from -0.137 (95% CI [-0.2641, -0.0103]) to -0.169 (95% CI [-0.3231, -0.0151]).[2]

Evaluation effort. Another way to test for systematic differences between blind and non-blind evaluation is to consider the amount of time a participant spent evaluating an idea. Potentially, innovation managers spend more time on non-blinded ideas in online evaluation; for example, if they come from the same unit or location. Conversely, they might spend less time on these ideas and simply use information about the proposer as a signal for idea quality. Model A5 in Table A3 uses *Effort*, the number of seconds spent on an idea, as the dependent variable and shows no systematic differences between the treatment conditions.[3]

Sample size and power. A more fundamental objection to our experimental design is that it is underpowered and therefore unable to detect the effects we seek. Per good practice, we calculated minimum detectable effect sizes prior to the experiment (Cohen, 1988), concluding that a sample of roughly 2,000 idea evaluations would allow us to detect effect sizes between 0.670 and 1.213.[4] Our estimates were much smaller (ranging from -0.0989 to 0.145), implying that we cannot expect them to become statistically significant in our sample. However, there is merit in interpreting estimation results beyond statistical significance. Even if a much larger sample would render our results statistically significant, the effect sizes would still be comparatively small. Consider the results for H2 in Model 2 from Table 2 —the largest effects we obtained. They imply that ideas from women were rated 0.401 points lower on a 7-point scale and, more importantly, that blind evaluation increases this rating by only 0.109 points. Even if these effects were estimated more precisely and rendered statistically significant, they still would be rather small (especially given that we cherry-picked the largest main effect for illustration).

Idea quality. Although we used real ideas from the company's idea management platform and screened them on basic criteria, we received open comments hinting at low quality. In total, we received 644 open comments (on 339 ideas), of which 205 (on 155 ideas) expressed that the idea was not clear enough and 137 (on 107 ideas) that the idea was not new. Low-quality ideas could impair our results because participants might come to take the evaluation less seriously and expend less effort. However, innovation managers spent on average 120 seconds evaluating an idea,

[1] We do not report all analyses that we outlined in the pre-analysis plan because some were rendered irrelevant by our null finding. To maximize transparency, we report the results of these analyses in Tables A4 and A5. In particular, we report all results on potential order effects in Table A5. We found no support for order effects as the negative effect of blind evaluation was only decreased by 0.0008 points (95% CI [-0.0105, 0.0090], Model A16) with every idea that was evaluated. We highlighted all exploratory analyses that were not in the pre-analysis plan with an asterisk.

[2] We also estimated *Next round* with a logistic regression. Results are shown in Table A4 (Models A9–A11).

[3] In our pre-analysis plan, we speculated that effort might be a mechanism driving differences between blind and non-blind evaluation. We provide additional analyses based on this idea in Table A4 (Models A12 and A13). One notable result is that innovation managers spent less time on each idea as the evaluation progressed (irrespective of treatment condition), implying that fatigue kicks in eventually.

[4] The effect sizes varied depending on the exact calculation; in particular, depending on the degree of correlation between evaluations from one participant. The calculations assumed a 7-point Likert scale with a mean of 3.5 and a common standard deviation of 1.5.

suggesting that they did consider the short idea descriptions quite carefully.[5] Another speculation would be that the effects of blinding only become visible above a certain quality threshold. Having too many low-quality ideas would then mask these effects. In Table A6, we report different models which exclude low-quality ideas. Although some coefficients prove unstable (especially those for female proposers), the results suggest that our null finding is not driven by low idea quality.

Distributional effects. Our experimental intervention could have distributional effects uncaptured by averages. For instance, the evaluator may like the idea proposer and give a higher score or dislike the proposer and give a lower score. Such behaviors would not necessarily produce different averages between the treatment conditions, as they could cancel each other out, but they would produce different distributions, inducing more extreme ratings in the non-blind evaluation and thus creating fatter tails. However, Figure A2 and Figures A4–A7 do not support this supposition. Moreover, Wilcoxon rank-sum tests failed to reject the null hypotheses that the two distributions are equal for *Overall score* ($p = 0.3727$) and all other outcomes (p-values between 0.1520 and 0.9191).

Demand effects. Given that innovation managers evaluated ideas in the blinded and the non-blinded condition, it is possible that an experimenter demand effect is at play. After a few evaluations, participants could suspect the blinding as a central part of the investigation and adjust their behavior accordingly. Were this the case, one would expect differences between evaluations made early in the random sequence (before participants suspected what the experiment was about) and late (when they might have become suspicious). To test this idea, we split the sample into early and late evaluations in two ways—the first versus the second half and the first 10 evaluations versus the rest. We report the results in Table A7. They show that the effect of blinding in the first half was very comparable to that in the second half. Comparing the first 10 evaluations to the rest shows that evaluators rated blinded ideas less favorably later, which speaks against demand effects, in which one would expect evaluators to become fairer and react *less* to the blinding once they had spotted it as the central research interest.

Alternative time thresholds. In our main analyses, we excluded evaluations that were completed in less than 8 seconds to ensure that evaluators did not just randomly click through the evaluation. At the same time, 8 seconds constitutes a relatively low bar, and quick evaluations may introduce noise in the data. We, therefore, considered alternative thresholds and excluded the evaluations that were in the bottom 5 percent (below 20.795 seconds) and 10 percent (below 27.126 seconds) in evaluation duration. We report the results in Table A8. The coefficients obtained with the alternative time thresholds are very similar to those in our original sample. The biggest change is in the main effect for female proposers, but even that remained statistically indistinguishable from zero in all samples. The null finding thus persisted.

Experimental hurdles. Finally, as often happens when conducting fieldwork in organizations, there were unforeseen problems. First, some innovation managers were omitted from the original invitation and six evaluators joined after the experiment had started (five of whom completed the evaluation). We show in Table A9 that excluding the latecomers does not change our results (Models A42–A44). Second, 44 of our participants did not rate an idea along all five dimensions (incomplete evaluations) and 118 evaluations came from participants who did not evaluate all 48 ideas (unfinished evaluations). In Table A9, we show that neither excluding incomplete evaluations nor excluding unfinished evaluations changes the results substantially (Models A45–A50). Finally, we introduced a small incentive as a reaction to low participation. We detail the introduction of the incentive in Table A9 and provide split-sample analyses for idea evaluations completed pre- and post-incentive (Models A51–A56). While there are some differences between the two groups, most coefficients resemble those of the main analysis and thus do not fundamentally challenge our null finding.

---

[5] The average given is different from that in Table A1 because here we exclude idea evaluations that took longer than 10 minutes, as they probably reflect breaks from the evaluation.

A5. Online experiment on Prolific

As mentioned in the main manuscript, we conducted an online experiment to assess the replicability of the null finding and address the above challenges more rigorously and comprehensively (data and code for the online experiment will be publicly available). Replicating the field experiment as closely as possible, we also addressed five challenges that remained intractable in the field experiment: (a) sample size and statistical power, (b) large number of ideas per evaluator, (c) idea quality, (d) experimental design, and (e) focus of the treatment.

First, the sample size in the field experiment may not give enough statistical power to detect small effect sizes. We had managed to recruit 38 evaluators—an arduous task due to the low number of qualified evaluators and the time commitment. Because participation was voluntary and we were limited in incentivizing evaluators, the sample size was smaller than we had hoped. We ensured sufficient statistical power for the online experiment by recruiting a much larger group of 1,543 evaluators.

Second, we asked innovation managers to evaluate 48 ideas in the field experiment resembling the time pressure often present in online idea evaluation. At the same time, the evaluators may have felt overburdened and, at some point, paid less attention to the evaluation task or taken it less seriously. In the online experiment, therefore, we assigned evaluators only 10 ideas. We also incentivized them with above-average compensation (see below).

Third, idea quality varied in the field experiment, which may have impaired evaluation. Recall that our field experiment used *actual* ideas from employees and there is usually variation in the quality of crowdsourced ideas (Piezunka & Dahlander, 2015). Evaluators may have paid little attention to the idea proposer because encountering low-quality ideas early in the experiment discouraged them or left them with little room to favor certain proposers. Therefore, for the online experiment, we took high-quality ideas that had already received favorable evaluations—specifically, crowdfunding ideas that reached their funding goal—and standardized them in terms of length and presentation.

Fourth, the field experiment used a within design and exposed evaluators to both blinded and non-blinded ideas, which may have introduced experimenter demand effects. The evaluators may have become sensitive to the blinding, suspected it as the central aspect of the evaluation task, and adjusted their behavior accordingly. If that was the case, a between design—in which evaluators are given either only blinded or only non-blinded ideas—would yield different results. We, therefore, ran two versions of the online experiment; one replicating the original within design and one using a between design. We also included four questions on the perceived awareness of the research hypothesis (PARH scale) to assess whether the observed effects are due to demand effects (Rubin, 2016).

Fifth, we could not separately blind the proposer's gender, unit, and location in the field experiment. This enabled us to test multiple hypotheses and we accounted for it in the regression analysis. But because blinding affected multiple facets of the proposer's identity, the treatment could have several—potentially countervailing—effects. Therefore, we went one step further in the online experiment and focused solely on the proposer's gender. Gender is the most meaningful aspect of the proposer's identity in the online setting, and gender biases are thought to operate more generally than the in/out-group biases underlying H3 and H4. In addition, gender is easily evoked online and thus made for the "cleanest" treatment possible.

Participants

We recruited participants on Prolific, a platform that helps researchers recruit from a pool of registered and screened participants. Due to its high data quality, Prolific is considered one of the most reliable such platforms (Peer, Brandimarte, Samat, & Acquisti, 2017). It allows researchers to sample participants based on demographic criteria; we recruited participants who (a) are located in the US or UK, (b) have completed at least technical or community college,[6] (c) are employed (full- or part-time), and (d) work as either a manager (junior, middle, or upper) or researcher.[7] We focused on participants from the US and UK to ensure that they can easily determine the gender of the names we provided (more below). We used the filters on education, employment status, and occupational role to recruit participants who (potentially) have some experience in evaluating ideas and deciding on them. Given the focus on gender, we recruited a gender-balanced sample.

To calculate the appropriate sample size, we performed power calculations and used information from the field experiment about effect sizes and intraclass correlations. Given the small effect sizes in the field experiment, we set for a minimum detectable effect (MDE) of 0.2. The resulting sample size thus provides us with enough statistical power to detect an effect of moving from, for example, 4.1 to 4.3 on a seven-point Likert scale. Any effect smaller than this seems hardly (economically) relevant. The power calculations revealed that we would need 1,308 participants (327 per condition in two experiments), but we expanded the data collection to 1540 participants (385 per condition). As usual in online studies, we used an attention check to ensure data quality (Oppenheimer, Meyvis, & Davidenko, 2009);[8] we excluded from our analysis the 4.5 percent of the participants who failed. The final sample had 1473 participants (49.0% female; average age = 35.97 years), yielding an effective MDE of 0.19. We provide descriptive statistics on the participants in the first column of Table A10. Most work full-time (86% vs. 14% part-time) in a management position (90% vs. 10% researchers), are based in the UK (65% vs. 35% in the US), and possess at least an undergraduate degree (87% vs. 13% with technical or community college). On average, participants have almost 15 years of working experience.

We obtained 10 evaluations from each participant, resulting in 14,730 (1473 evaluators * 10 ideas). As in the field experiment, we excluded 67 evaluations that were answered in less than eight seconds. The final sample thus included 14,663. On average, participants spent 548 seconds (9.14 minutes) on the evaluation and received 8.04 pounds per hour (11.10 dollars) in compensation—more than the average compensation on Prolific.

Evaluation Task

In line with our focus on crowdsourced ideas in the technology sector, we asked each participant to evaluate 10 "ideas for apps for mobile phones." The evaluation setup closely resembled that of the field experiment, although we adjusted it for a non-corporate setting. We used the same survey as in the field experiment and replicated its design, but removed the corporate design of our industrial partner. We illustrate the evaluation screen in Figure 1 and the survey flow in Figure A1.

To ensure that ideas are high-quality and comparable, we chose 10 projects from Kickstarter's app category that had reached their funding goal between May 1, 2013 and April 2, 2021. We rewrote them in a standardized format, describing the problem the application addresses and the solution it offers in about 125 words. We list the ideas, their descriptions, and supplemental information in Table A11.

Treatment Conditions

---

[6] This includes participants who have an undergraduate (BA, BSc, or other), graduate (MA, MSc, or other), or doctoral (PhD or other) degree.

[7] To ensure that the online experiment worked as intended, we pretested it with 32 participants. The pretest was identical to the online experiment, but also asked participants to report any issues with the experiment. We used the pretest to see how much time participants needed to evaluate the ideas and to determine a fair compensation. We excluded participants in the pretest from the online experiment.

[8] The attention check was an additional question randomly assigned to one of the ideas. We followed best practices and Prolific's Attention and Comprehension Check Policy to design a fair attention check (for more information, see https://researcher-help.prolific.co/hc/en-gb/articles/360009223553-Prolific-s-Attention-and-Comprehension-Check-Policy). We wrote: "It is important that you pay attention. To show that you did, please select 'Somewhat disagree' when asked about it." Then we asked: "Based on the text above, what have you been asked to answer?" Participants were given the statement "It is important to pay attention" and a five-point scale ("Strongly disagree" to "Strongly agree").

The online experiment focused on blinding the idea proposer's gender. The effect of gender was strongest in the field experiment and gender is meaningful in the online context. It is more meaningful and less abstract than imagining organizational structures, which would be needed to test effects of unit and location (H3 and H4). Compared to blinding multiple pieces of information simultaneously, focusing on gender clarifies the treatment—that is, exactly what information is withheld.

Like the field experiment, the online experiment had two conditions: *blind evaluation*, in which participants received no information about the idea proposer and *non-blind evaluation*, in which participants received the (fictional) proposer's name (for example, "This idea was submitted by: Anne"). Consistent with our focus on gender, the non-blind condition provided only the proposer's name (in the field experiment, it also provided unit and location). We chose five clearly female names (Emily, Anne, Kristen, Allison, and Sarah) and five clearly male names (Greg, Todd, Brad, Matthew, and Neil) and randomly allocated them to the 10 ideas. These names are of similar socioeconomic status and evoke the same race (Bertrand and Mullainathan, 2004), holding constant other factors that might affect evaluation.[9] The proposer's name thus mainly signals his or her gender. To attract attention, the names were printed in bold and highlighted in yellow and we found in pretests that participants were paying attention to them.[10]

We ran two versions of the online experiment, which differed in how we assigned the treatment conditions. First, we replicated the field experiment and used a *within-subject design*: participants evaluated ideas under both conditions, randomly blinded for the proposer of *some ideas*, but not of others. Second, we used a *between-subjects design* to investigate whether the experimental design changes the results. Participants were randomly assigned to evaluate *all ideas* in either the non-blind or blind condition.

Randomization

In a first step, each participant was randomly assigned to an experimental design: within or between. For the between design, he or she was then assigned randomly to either blind or non-blind evaluation. For the within design, information about the idea proposer was withheld at random for some ideas. In both cases, we randomized the order of the ideas to control for order and sequence effects (Bian et al., 2021; Criscuolo et al., 2021). We also randomly assigned (fictional) proposers to the ideas. In the field experiment, we knew the *real* identity of the proposer whom we randomly blinded. In the online experiment, however, we randomly assigned names to the ideas. Since all participants evaluate all 10 ideas, we did not need to pick ideas from a larger pool as in the field experiment. However, we randomly added the attention check question to one of the ideas.

Table A10 provides balance checks for the experimental designs (within vs. between; Columns 2–4) and the experimental conditions in the between design (non-blind vs. blind; Columns 5–7). There are almost no significant differences in the groups' demographics, as indicated by low t-statistics and high p-values.[11] The groups are thus well balanced, and we find that randomization—for both experimental design and treatment condition—worked as intended. This makes it unnecessary to add demographic controls for the participants in the estimations and implies that mean comparisons among the groups are informative.

Variables

---

[9] To ensure that potential differences in the inferred socio-economic status do not affect our results, we also ran a model in which we controlled for socioeconomic status (proxied by mother's education associated with the names, see Bertrand & Mullainathan, 2004). The results are robust to this inclusion and available upon request.

[10] We did informal pre-tests with ten people from one of our home institutions. We presented the participants the evaluation screen and asked them to evaluate a few ideas. In a short follow-up interview, we asked them about their general experience and whether they had recognized the idea proposer's name.

[11] The exception is that there are significantly more participants with a graduate degree in the within design than in the between design (p = 0.0219). However, there is little reason to believe that this one imbalance causes any systematic bias in our results.

Dependent variables. We used the same dependent variables (with the same scales) as in the field experiment; *Overall score*, *Desirability*, *Feasibility*, *Viability*, and *Next round*. As in the field experiment, *Overall score* is the main dependent variable. We adjusted the wording of the questions slightly to fit the online context. The items were defined as follows:

Desirability: The degree to which a solution to a problem addresses someone's needs.

Feasibility: The degree to which a solution is possible and suitable to implement.

Viability: The degree to which an idea makes business sense.

Treatment variable. As in the field experiment, we used the dummy variable *Blind* to indicate that a participant evaluated an idea in the blind condition.

Independent variables. Because we randomly assigned male and female names to the ideas, ideas and idea proposers are not fixed pairs (as they were in the field experiment) and we do not need to specify an interaction to analyze the effect of blinding. Instead, we compare assigning male and female names to the blind condition, captured by the indicator variables *Female name* and *Male name*.

Additional variables. To investigate potential experimenter demand effects, we included the established perceived awareness of research hypothesis scale (PARH scale; Rubin, 2016) in the exit survey. This is a self-report measure with four items measured on a seven-point Likert scale. To measure the degree to which participants believed they had an understanding of the research hypothesis, we included the established perceived awareness of research hypothesis scale (PARH scale) by Rubin (2016) in the exit survey of the online experiment. The scale has four items and participants were asked to which degree they agreed or disagreed with the following statements:

1. I knew what the researchers were investigating in this research.
2. I wasn't sure what the researchers were trying to demonstrate in this research.
3. I had a good idea about what the hypotheses were in this research.
4. I was unclear about exactly what the researchers were aiming to prove in this research.

Participants responded to these questions on a seven-point Likert scale ranging from strongly disagree (1) to strongly agree (7). Items two and four were reverse coded and individual item scores were averages in a single overall PARH score. High PARH scores mean that participants believe they had a good idea of the research hypothesis. We also asked an open-ended question ("What do you think this study was about?") to elicit whether they knew what we sought to test.

Results

Table A12 provides an overview of the variables and their correlations for the two versions of the online experiment. In general, the numbers are very comparable across designs. In both experiments, exactly 50 percent of the ideas were rated in a blind evaluation and 25.1 and 24.9 percent of the ideas were assigned a female name. These numbers confirm that random assignment worked technically. Compared to the field experiment, the ideas received higher scores on all dimensions (*Overall score*, *Next round*, *Desirability*, *Feasibility*, and *Viability*), showing that the ideas in the online experiment were perceived to be of higher quality than those in the field experiment. For *Overall score*, the ideas in the online experiment received an average of 4.64 (within design) and 4.58 (between design), whereas the ideas in the field experiment received an average of 3.32. Similar to the field experiment, there was relatively little consensus among the evaluators who agreed on the rating in 19.28% of the cases. The intraclass correlation coefficient for ideas was 0.0951 (two-way random effects model).

We took the same steps to analyze the data as in the field experiment. First, we conducted mean comparisons of the non-blind and blind conditions for both the within and between designs. Table A13 shows that the differences, as in the field experiment, were small (the highest difference being 0.0580) and t-tests failed to reject that blind and non-blind evaluations produce the same mean outcomes. We find that *Overall score* was slightly higher in non-blind evaluation for the within design, with a marginally statistically significant difference ($p = 0.10$). The mean comparisons also support the appropriateness of the within design: Table A13 shows that the average scores in the blind condition are very similar for the two experimental designs. Assuming that blind evaluation in the between design reflects a baseline of the idea's "true" value (because participants evaluate ideas in the absence of any proposer information), this result shows that the blind condition in the within design does a good job of replicating this baseline. That, in turn, validates the within design used in the field experiment. Next, we compared the distribution of *Overall score* graphically in Figure 3. The four distributions look almost identical and show no notable differences between blind and non-blind evaluations and between the two experimental designs. Again, Wilcoxon rank-sum tests failed to reject that the distributions are equal (p = 0.0769 for within design, p = 0.8251 for between design).[12]

Finally, we ran OLS estimations to test the effect of blinding in a regression setup. We used the treatment indicator *Blind* to test the main effect of blinding (H1). To test the effect of assigning a female or male name to the idea (H2), we used the dummy variables *Female name* and *Male name*. We use the blind condition as a baseline and compare the effect of assigning a female or male name to this condition. Given that we randomly *assigned* the proposer names to the ideas (and that proposers and ideas are not fixed pairs as in the field experiment), we did not specify an interaction, but the main effects of *Female name* and *Male name* give the difference between presenting female and male proposers and the blind condition. As in the field experiment, we included fixed effects for the idea and for the display order. Note that including fixed effects for the evaluator is possible only in the within design because there is no within-evaluator variation in blinding for the between design.

We report the estimation results in Table A14. In the within design, we find that ideas get slightly lower ratings in blind evaluation. However, the effect is small (point estimate of -0.0629; 95% CI [-0.1257, -0.0001]) and hardly economically meaningful. Its statistical significance is largely the result of a larger sample. In the between design, we find no difference between blind and non-blind evaluation (point estimate of 0.0124; 95% CI [-0.0549, 0.0798]). In neither design do we find differences in the evaluation scores of ideas proposed by women or men, compared to those in the blind condition. Wald tests for the equality of the coefficients of *Female name* and *Male name* show no differences. This pattern holds for all dependent variables—we find that assigning female or male names does not affect idea evaluation.[13]

We thus found some support for H1 that ideas are rated lower in blind evaluation. However, the effect is small and statistically distinguishable from zero only in the within design. We found no evidence for H2 that the gender of the idea proposer matters for evaluation scores. Overall, the results from the online experiment confirm the findings from the field experiment. Independently of the experimental design, we find no meaningful differences between blind and non-blind evaluation. Moreover, assigning female or male names to the ideas did not change evaluation scores compared to the baseline of blind evaluation, which is in line with the field experiment, in which blinding had no significant effect on female proposers' evaluation scores. Table A16 summarizes the key estimates across the field and online experiment and underlines the consistency of their results.

Post-hoc analyses

Although the between and within designs yielded similar results, suggesting that experimenter demand effects did not drive our results, we further scrutinized such effects. We also analyzed whether our results were shaped by evaluator gender and tested alternative time thresholds.

Ruling out demand effects. We investigated potential demand effects through the PARH score to analyze whether participants believed they understood our hypotheses. Participants had an average PARH score of 4.18 in the within design and 4.20 in the between design. These averages are close to the midpoint of the seven-point scale and do not suggest that participants had a clear idea of our hypotheses. More importantly, the difference between the two designs is tiny and is

[12] We produced distribution graphs for the alternative dependent variables *Next round*, *Desirability*, *Feasibility*, and *Viability* and show them in Figures A6 to A9. They also exhibit no notable differences between the two conditions, which is supported by Wilcoxon rank-sum tests (p-values between 0.2564 and 0.6819).

[13] We report the results for the other outcomes (*Next round*, *Desirability*, *Feasibility*, *Viability*) in Table A15.

statistically indistinguishable from zero (p = 0.8348), which shows that participants did not perceive the hypotheses as more obvious in the within design. Supporting this, participants in the blind condition of the between design, who by design cannot know the hypotheses, had a PARH score statistically indistinguishable from that of participants in the non-blind condition (4.19 vs. 4.20; p = 0.9145). This shows that even if participants thought they understood the hypotheses reasonably well, they thought so irrespective of whether they could or could not actually know. We also reviewed the participants' open-text answers on what they thought the study was about. Out of the 268 cases in which participants were confident they knew the research hypothesis (defined by a PARH score greater than or equal to 6), only 3 (1.12%) were correct. Overall, these additional analyses prompt little concern for experimenter demand effects.

Evaluator gender. Female and male evaluators may react differently to proposers of the same and different gender. We tested this by including a dummy variable *Female evaluator* and interacting it with *Female name* and *Male name*. We report the results in Table A14. We found no different reaction to female and male names among female and male evaluators. However, female evaluators evaluated the ideas more favorably in general.

Alternative time thresholds. As in the field experiment, we again considered alternative time thresholds and excluded the evaluations that were in the bottom 5 percent (below 13.439 seconds) and 10 percent (below 16.754 seconds) in evaluation duration. We report the results in Table A17. Again, the coefficients only vary slightly across the samples and the null finding is robust to using alternative time thresholds.

Online experiment conclusion

The online experiment circumvented many challenges that were intractable in the field experiment. We ensured a sufficiently large sample and statistical power, higher and more constant idea quality, and fewer ideas per evaluator. We still found no differences between blind and non-blind evaluation, thereby supporting the null finding of the field experiment. Another advantage of the online experiment is that it helps rule out that the field experiment's null finding resulted from its design. We found no differences between the within and between designs and replicated the null finding with a "cleaner" treatment that focused only on the proposer's gender.

A6. Vignette study of expected effect sizes

The field and online experiment results suggest that blinding has little or no effect on idea evaluation. We found no differences between blind and non-blind evaluations. The null findings contradict our theorizing, as outlined in the pre-analysis plan and main text. We conducted a final vignette study to critically review our theoretical priors and investigate common beliefs about biases in idea evaluation. We presented participants the experimental design of the online experiment and asked them to estimate the gender difference in evaluation scores.

Participants

We again recruited participants on Prolific. As we were looking for highly educated and independent participants, we recruited 213 participants holding or pursuing a doctorate degree who reside in the US or UK and have not participated in the online experiment (58.7% female; average age = 29.92 years). On average, participants took 112 seconds (1.93 minutes) and we paid them 20.46 pounds per hour (28.25 dollars).[14] We also incentivized them with a bonus payment of 1 pound for the 25% of the most accurate guesses.

Estimation task

We gave participants complete information about our online experiment and its design and then asked them to guess the average evaluation scores for ideas proposed by men and women. We used the following instructions to ask the participants about their estimations: "In a previous study we asked evaluators to rate ten new app ideas. The name of the idea proposer was shared with the evaluator. As we wanted to explore the impact of gender on how new app ideas are evaluated, we

[14] We eliminated one obvious outlier when calculating the average duration.

randomly assigned either a male (Greg, Todd, Brad, Matthew and Neil) and a female (Emily, Anne, Kristen, Allison and Sarah) name to each idea. The example below shows how new app ideas were presented to evaluators." We then provided an example of the evaluation screen in the online experiment (similar to the right side of Figure 1). The estimation screen is shown in Figure A11. This screen automatically calculated the difference as the participants moved the slider across the continuous scales.

Results

On average, participants estimated that the evalaution scores would be 4.59 (S.D. = 0.90) for ideas with a female name and 5.24 (S.D. = 0.88) for ideas with a male name. Participants thought the difference would be 0.65 in favor of men.[15] We provide the whole distribution of the estimates in Figure A11. We also analyzed whether men and women made different estimations and found that women estimated the difference to be significantly bigger (0.89 for women vs. 0.30 for men).

Vignette experiment conclusion

The independent participants' expectations about gender biases in idea evaluation were close to our own priors. However, they stand in sharp contrast to our findings from the online experiment, in which we discovered no difference in evaluation scores for women and men. Even when independent participants knew our exact design and measures, they overestimated the difference.

REFERENCES

Baumeister R. F., Bratslavsky E., Muraven M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology 74*(5): 1252-1265.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review 94*(4), 991-1013.

Bian, J., Greenberg, J., Li, J., & Wang, Y. (2021). Good to go first? Position effects in expert evaluation of early-stage ventures. *Management Science 68*(1), 300-315.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2021). The sequence effect in panel decisions: Evidence from the evaluation of research and development projects. *Organization Science 32*(4), 987-1008.

de Bruin, W. B. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica 118*(3), 245-260.

de Bruin, W. B., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes 92*(1–2), 91-101.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison of preferences. *Journal of Experimental Social Psychology 25*(2), 121-141.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology 45*(4), 867-872.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology 70*, 153-163.

Rubin, M. (2016). The perceived awareness of the research hypothesis scale: Assessing the influence of demand characteristics. *Figshare 61*(0), 1-6.

[15] Remember that the MDE for the field experiment was 0.670 (under favorable, but realistic assumptions), which is close to the difference of 0.65 estimated in the vignette study.

FIGURES

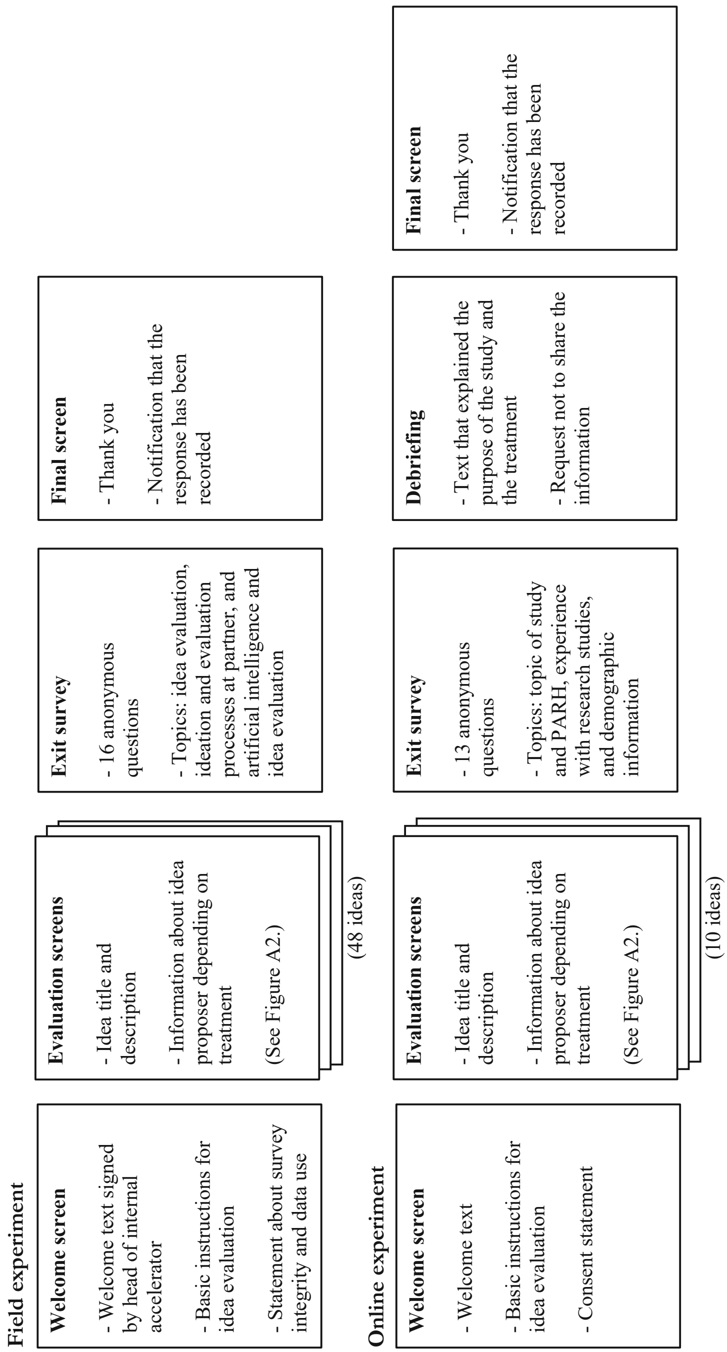


FIGURE A1 Survey flow of field experiment and online experiment

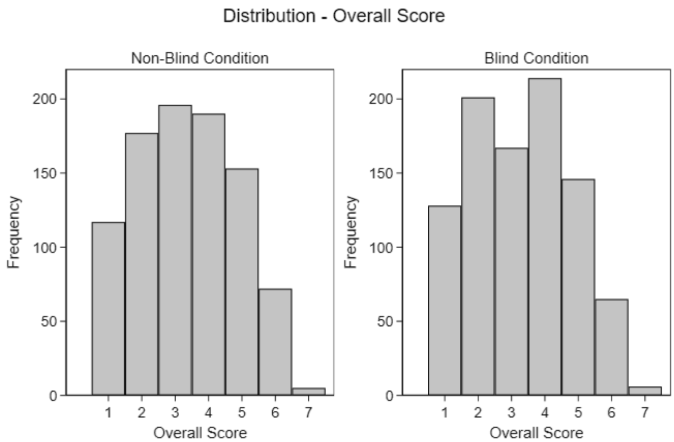


FIGURE A2 Field experiment: Graphic comparison between treatment conditions

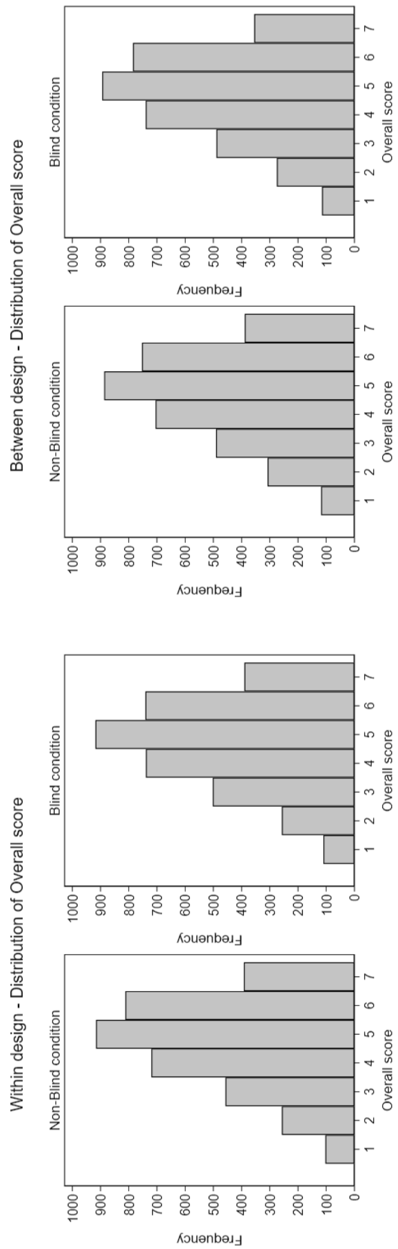


FIGURE A3 Online experiment: Graphic comparison between treatment conditions

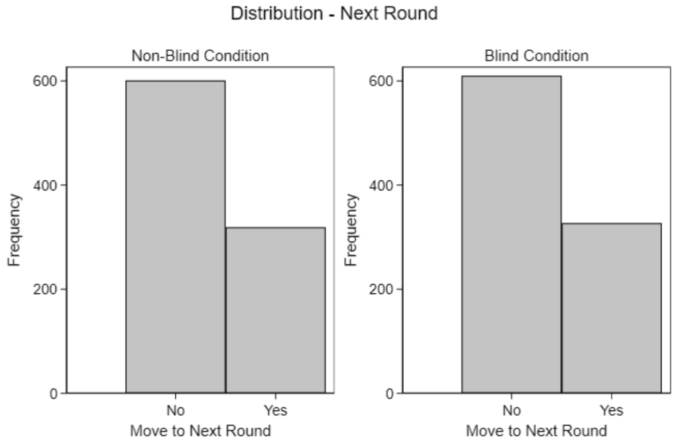


FIGURE A4 Field experiment: Graphic comparison between treatment conditions (Next round)

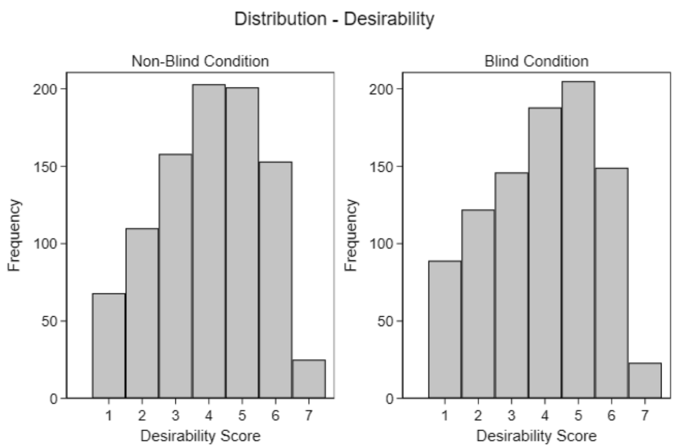


FIGURE A5 Field experiment: Graphic comparison between treatment conditions (Desirability)

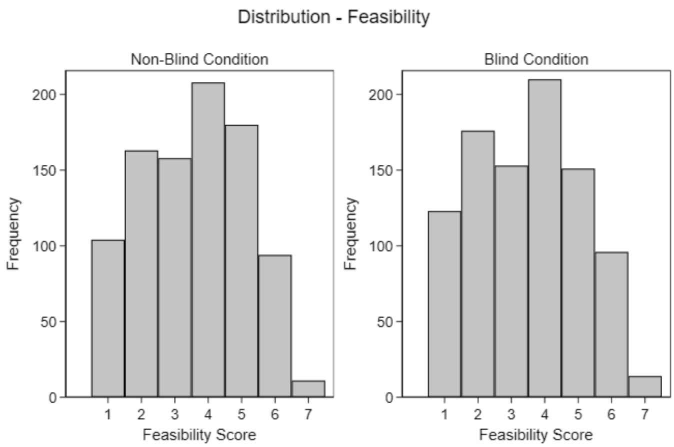


FIGURE A6 Field experiment: Graphic comparison between treatment conditions (Feasibility)

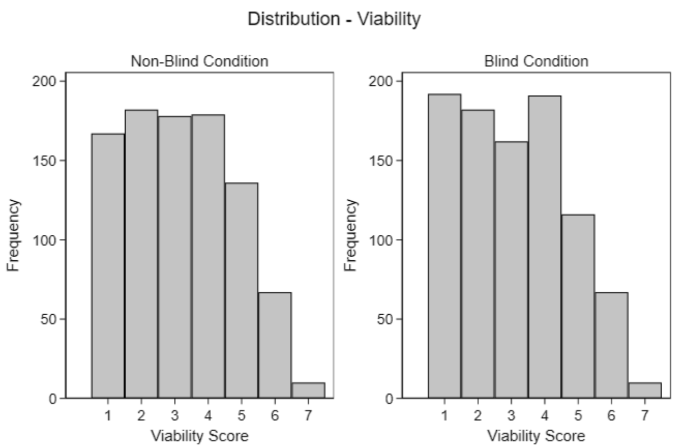


FIGURE A7 Field experiment: Graphic comparison between treatment conditions (Viability)

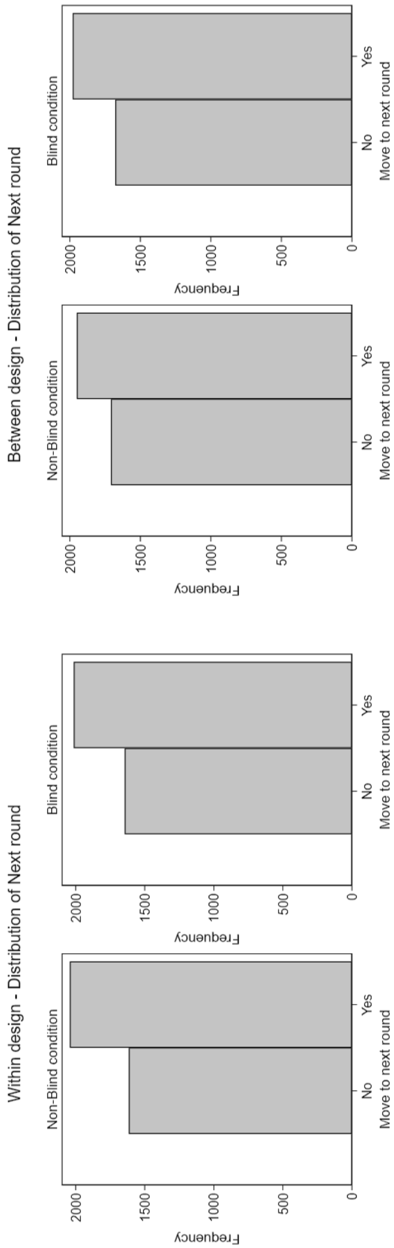


FIGURE A8 Online experiment: Graphic comparison between treatment conditions (Next round)

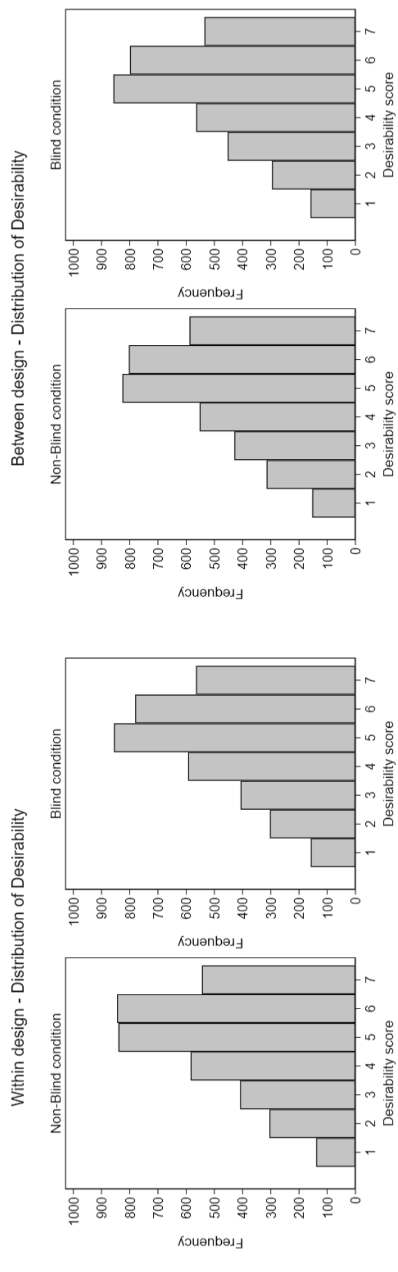


FIGURE A9 Online experiment: Graphic comparison between treatment conditions (Desirability)

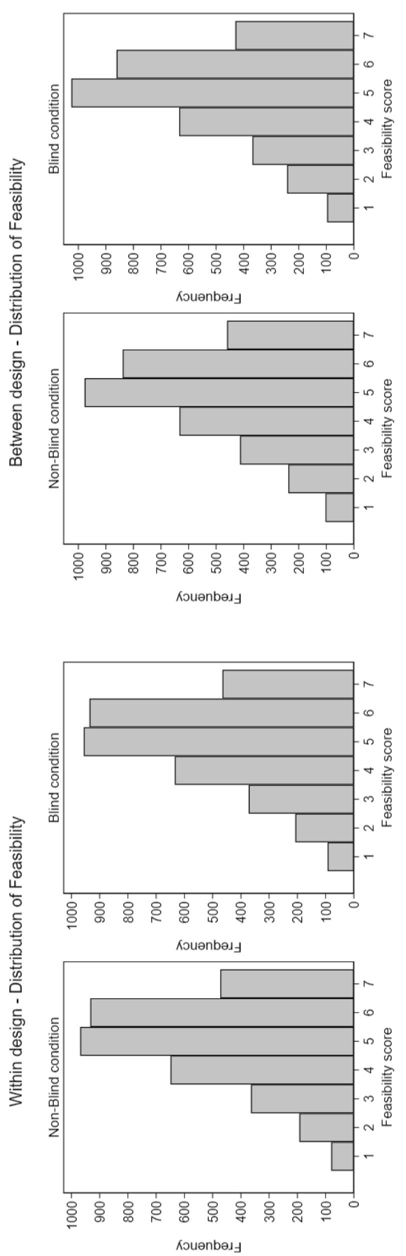


FIGURE A10 Online experiment: Graphic comparison between treatment conditions (Feasibility)

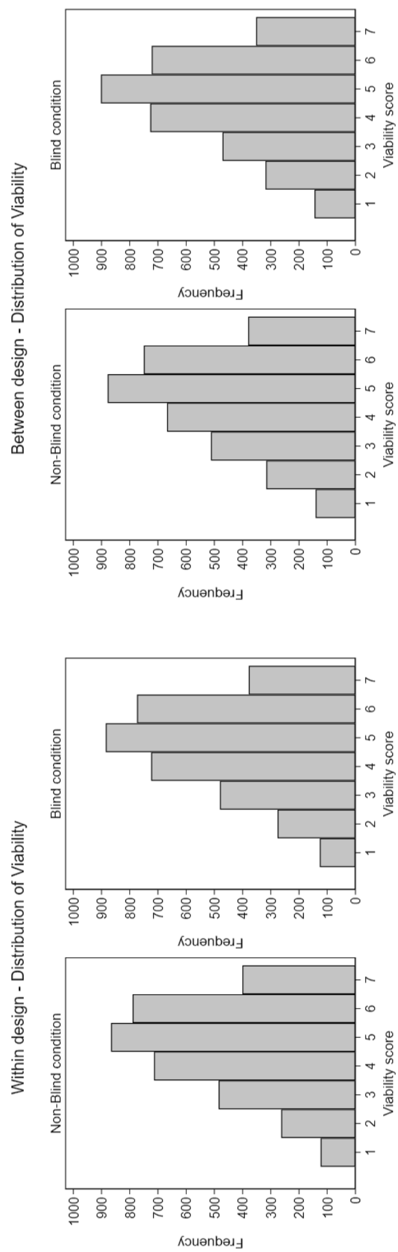


FIGURE A11 Online experiment: Graphic comparison between treatment conditions (Viability)

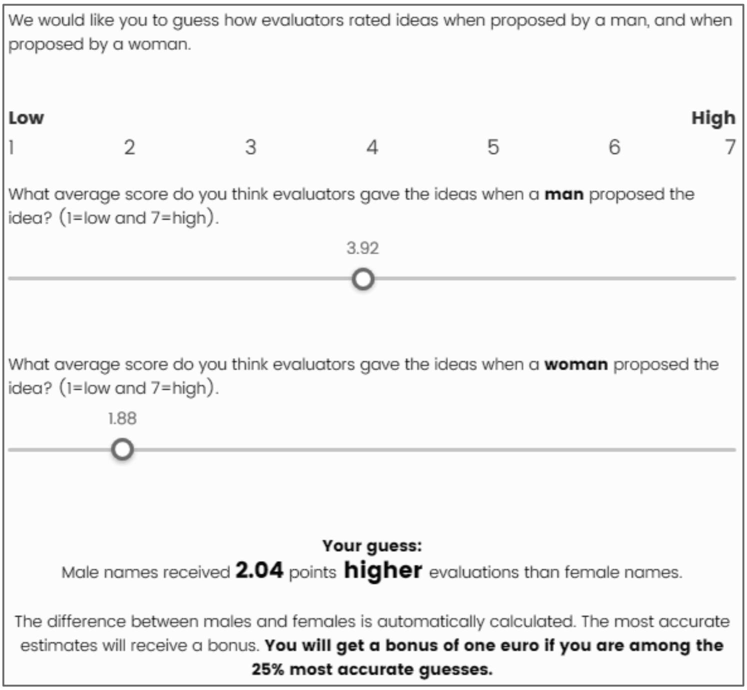


FIGURE A12 Vignette study: Screenshot of the estimation screen

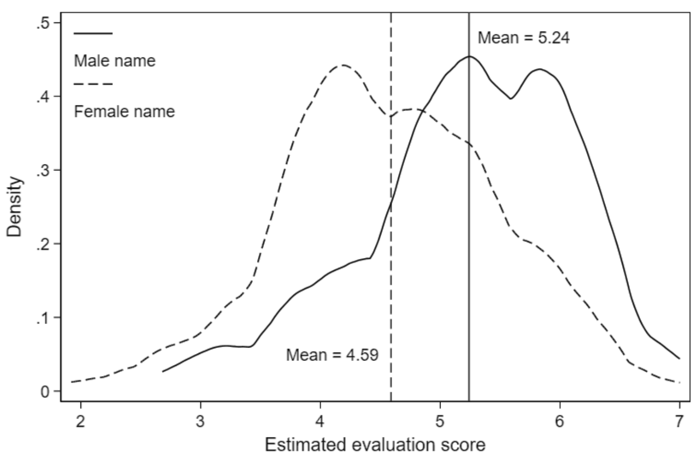


FIGURE A13 Vignette study: Kernel density of estimated evaluation scores

TABLES

TABLE A1 Field experiment: Descriptive statistics

	Mean	S.D.	Min	Max	1	2	3	4	5	6	7	8	9
1 Blind	0.502	0.500	0	1									
2 Overall score	3.321	1.509	1	7	-0.02								
3 Next round	0.348	0.476	0	1	0.00	0.70							
4 Desirability	3.954	1.596	1	7	-0.03	0.75	0.57						
5 Feasibility	3.520	1.581	1	7	-0.03	0.77	0.61	0.65					
6 Viability	3.149	1.593	1	7	-0.03	0.85	0.68	0.64	0.78				
7 Duration	590.103	4853.062	8.315	153563.65	0.04	0.04	0.04	0.04	0.03	0.03			
8 Female proposer	0.167	0.373	0	1	0.01	-0.06	-0.04	-0.06	-0.07	-0.06	0.00		
9 Same unit	0.129	0.335	0	1	0.01	0.06	0.03	0.07	0.03	0.06	0.02	-0.01	
10 Same location	0.095	0.293	0	1	0.03	-0.04	-0.06	-0.03	-0.06	-0.05	0.01	-0.00	0.14

TABLE A2 Field experiment: Alternative dependent variables - Mean comparisons of treatment conditions

	Non-blind condition (910 evaluations)	Blind condition (927 evaluations)	Diff.	t-stat.	p-value
Next round	0.347	0.349	-0.00225	-0.10	0.92
Desirability	4	3.908	0.0922	1.24	0.22
Feasibility	3.570	3.470	0.0995	1.35	0.18
Viability	3.192	3.107	0.0850	1.14	0.25

TABLE A3 Field experiment: Alternative dependent variables - Regression analyses

	(A1) Next round	(A2) Desirability	(A3) Feasibility	(A4) Viability	(A5) Effort
H1: Blind	-0.0234	-0.137	-0.169	-0.151	-1.829
	(0.262)	(0.0347)	(0.0321)	(0.0408)	(0.750)
Female proposer	0.0269	-0.644	0.265	-0.460	-3.081
	(0.911)	(0.521)	(0.775)	(0.539)	(0.922)
Same subunit	0.0595	0.149	0.0769	0.204	16.09
	(0.247)	(0.297)	(0.633)	(0.148)	(0.159)
Same location	-0.149	-0.0861	-0.395	-0.338	-4.303
	(0.0268)	(0.672)	(0.0142)	(0.0436)	(0.740)
H2: Blind x Female proposer	0.0729	0.0759	0.0790	0.114	-0.0205
	(0.223)	(0.669)	(0.675)	(0.471)	(0.999)
H3: Blind x Same subunit	-0.0446	-0.0273	0.123	0.0778	4.327
	(0.497)	(0.902)	(0.624)	(0.719)	(0.839)
H4: Blind x Same location	0.0830	0.179	0.175	0.130	0.220
	(0.246)	(0.382)	(0.349)	(0.465)	(0.991)
Constant	0.198	4.249	3.115	3.195	176.8
	(0.0992)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes
N	1857	1840	1841	1839	1789
R^2	0.3507	0.4010	0.3755	0.4028	0.3490

Notes. Dependent variables are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Model A5 excluded individual idea evaluations that took longer than 10 minutes. The sample size varied between models because innovation managers did not always evaluate the ideas on all dimensions.

TABLE A4 Field experiment: Preregistered analyses

	Excluding regular users of idea management system (DV: Overall score)			Logistic regression (DV: Next round)			Effort mechanism (DV: Effort)		Split sample – Pre and post incentive (DV: Effort)	
	(A6)	**(A7)**	**(A8)**	**(A9)**	**(A10)**	**(A11)**	**(A12)**	**(A13)**	**(A14) Pre**	**(A15) Post**
H1: Blind	0.0578	0.201	0.277	-0.0909	0.101	-0.252	-1.178	5.192	2.503	-6.080
	(0.606)	(0.419)	(0.0924)	(0.489)	(0.724)	(0.148)	(0.803)	(0.579)	(0.722)	(0.647)
Order		0.00117			-0.00176			-1.462		
		(0.885)			(0.840)			(0.000)		
HA1: Blind x Order		-0.00561			-0.00991			-0.327		
		(0.493)			(0.365)			(0.377)		
Female proposer			0.533			-0.195				
			(0.641)			(0.902)				
Same subunit			0.236			0.593				
			(0.504)			(0.205)				
Same location			0.185			-1.466				
			(0.659)			(0.00292)				
H2: Blind x Female proposer			-0.594			0.560				
			(0.0656)			(0.233)				
H3: Blind x Same subunit			-0.422			-0.446				
			(0.311)			(0.384)				
H4: Blind x Same location			-0.340			1.062				
			(0.482)			(0.0626)				
Constant	2.270	2.252	2.183	-1.071	-0.822	-0.548	177.1	111.4	179.9	202.9
	(0.001)	(0.000)	(0.001)	(0.357)	(0.444)	(0.661)	(0.000)	(0.000)	(0.001)	(0.010)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
N	713	713	713	1428	1428	1428	1789	1789	996	793
R^2	0.6948	0.6444	0.7004				0.3466	0.3181	0.5180	0.6411
Pseudo R^2				0.3185	0.2817	0.3261				

Notes. Dependent variables, sample restrictions and/or estimation techniques are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models A6 to A8 excluded all evaluators who indicated that they have indicated that they have logged in to the idea management system 6 times or more in the past six months. Models A12 to A15 excluded individual idea evaluations that took longer than 10 minutes.

TABLE A5 Field experiment: Preregistered analyses on order effects

	Order effects (DV: Overall score)	Non-linear order effects (DV: Overall score)	
	(A16)	(A17)	(A18)
H1: Blind	-0.0744	-0.240	-0.0723
	(0.562)	(0.172)	(0.572)
Order	-0.00616	-0.0142	
	(0.117)	(0.296)	
HA1: Blind x Order	-0.000782	0.0199	
	(0.873)	(0.188)	
$Order^2$		0.000169	
		(0.526)	
Blind x $Order^2$		-0.000430	
		(0.154)	
Second 10			-0.0581
			(0.680)
Third 10			0.0178
			(0.913)
Fourth 10			-0.266
			(0.0758)
Fifth 10			-0.122
			(0.500)
Blind x Second 10			0.0949
			(0.540)
Blind x Third 10			-0.141
			(0.493)
Blind x Fourth 10			0.104
			(0.620)
Blind x Fifth 10			-0.211
			(0.349)
Constant	2.493	2.549	2.436
	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes
Order fixed effects	No	No	No
N	1837	1837	1837
R^2	0.3926	0.3933	0.3948

Notes. Dependent variables are given in column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. *Order* is a count variable that recorded at what point in the evaluation process an idea was evaluated. For example, *Order* would be 13 if innovation manager j evaluated idea i as the 13th idea. Model A17 contained additional squared terms for *Order*, and Model A18 contained a set of indicator variables grouping individual idea evaluations by order in batches of ten (with the first ten idea evaluations as the baseline).

TABLE A6 Field experiment: Idea quality

	Excluding unclear ideas (DV: Overall score)			Excluding already existing ideas (DV: Overall score)			Excluding low-ranked ideas (DV: Overall score)		
	(A19)*	**(A20)***	**(A21)***	**(A22)***	**(A23)***	**(A24)***	**(A25)***	**(A26)***	**(A27)***
H1: Blind	-0.142	-0.130	-0.174	-0.0724	-0.0780	-0.144	-0.101	-0.0839	-0.129
	(0.0554)	(0.406)	(0.0716)	(0.357)	(0.556)	(0.111)	(0.138)	(0.558)	(0.111)
Order		-0.00405			-0.00653			-0.00699	
		(0.406)			(0.0916)			(0.105)	
HA1: Blind x Order		-0.000983			0.0000484			-0.000502	
		(0.877)			(0.993)			(0.926)	
Female proposer			2.285			-0.505			-0.363
			(0.00258)			(0.506)			(0.608)
Same subunit			0.165			0.187			0.216
			(0.376)			(0.296)			(0.201)
Same location			-0.114			-0.255			-0.231
			(0.585)			(0.271)			(0.249)
H2: Blind x Female proposer			0.0701			0.314			0.109
			(0.743)			(0.0422)			(0.472)
H3: Blind x Same subunit			0.0423			0.113			-0.00579
			(0.871)			(0.671)			(0.980)
H4: Blind x Same location			0.211			0.115			0.105
			(0.441)			(0.609)			(0.642)
Constant	2.608	2.489	2.606	2.552	2.449	2.588	2.613	2.513	2.619
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
N	1105	1105	1105	1374	1374	1374	1707	1707	1707
R^2	0.3850	0.3537	0.3871	0.4094	0.3915	0.4148	0.3442	0.3303	0.3473

Notes. Dependent variables and sample restrictions are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models highlighted with an asterisk were not planned in the preregistration. Models A19 to A21 excluded all ideas where at least one open comment indicated that the idea was not clear enough. Models A22 to A24 excluded all ideas where at least one open comment indicated that the idea already existed or was not new. Models A25 to A27 excluded all ideas where the idea fixed effect in Model 2 was estimated to be negative at the 5% significance level.

TABLE A7 Field experiment: Demand effects

	First vs. second half evaluations (DV: Overall score)				**First ten evaluations vs. rest (DV: Overall score)**			
	First 24 (A28)*	**Last 24 (A29)***	**First 24 (A30)***	**Last 24 (A31)***	**First 10 (A32)***	**Rest (A33)***	**First 10 (A34)***	**Rest (A35)***
H1: Blind	-0.109	-0.179	-0.123	-0.116	0.0366	-0.168	-0.0758	-0.151
	(0.286)	(0.271)	(0.355)	(0.560)	(0.879)	(0.0241)	(0.780)	(0.117)
Female proposer			0.881	-2.229			2.480	-1.862
			(0.396)	(0.0797)			(0.0372)	(0.0379)
Same subunit			0.0136	0.322			-0.116	0.0331
			(0.957)	(0.265)			(0.833)	(0.841)
Same location			-0.146	0.167			-0.0576	-0.146
			(0.659)	(0.681)			(0.913)	(0.598)
H2: Blind x Female proposer			-0.106	0.0415			-0.0371	0.000177
			(0.712)	(0.868)			(0.944)	(0.999)
H3: Blind x Same subunit			0.151	-0.0664			0.268	0.0731
			(0.617)	(0.876)			(0.705)	(0.789)
H4: Blind x Same location			0.171	-0.562			0.571	-0.220
			(0.680)	(0.294)			(0.426)	(0.477)
Constant	2.319	2.721	2.298	2.671	0.866	3.349	0.885	3.356
	(0.000)	(0.0102)	(0.000)	(0.0136)	(0.278)	(0.000)	(0.322)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	992	845	992	845	429	1408	429	1408
R^2	0.5419	0.5722	0.5429	0.5773	0.8227	0.4691	0.8249	0.4717

Notes. Dependent variables and sample restrictions are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models highlighted with an asterisk were not planned in the preregistration.

TABLE A8 Field experiment: Alternative time thresholds

	8 seconds threshold (original)		5th percentile threshold		10th percentile threshold	
	(A36)	**(A37)**	**(A38)***	**(A39)***	**(A40)***	**(A41)***
	Overall score	**Overall score**	**Overall score**	**Overall score**	**Overall score**	**Overall score**
H1: Blind	-0.0989	-0.133	-0.0975	-0.133	-0.0587	-0.0916
	(0.119)	(0.0770)	(0.168)	(0.0921)	(0.448)	(0.279)
Female proposer		-0.401		-0.0963		-0.0534
		(0.583)		(0.921)		(0.956)
Same subunit		0.170		0.176		0.136
		(0.280)		(0.257)		(0.404)
Same location		-0.235		-0.255		-0.225
		(0.211)		(0.195)		(0.265)
H2: Blind x Female proposer		0.109		0.0856		0.0677
		(0.462)		(0.585)		(0.691)
H3: Blind x Same subunit		0.0258		0.0436		0.0475
		(0.907)		(0.844)		(0.835)
H4: Blind x Same location		0.145		0.187		0.171
		(0.463)		(0.373)		(0.424)
Constant	2.696	2.703	2.619	2.622	2.646	2.650
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,837	1,837	1,766	1,766	1,674	1,674
R^2	0.4049	0.4074	0.4081	0.4108	0.4145	0.4164

Notes. Dependent variables are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models highlighted with an asterisk were not planned in the preregistration.

TABLE A9 Field experiment: Experimental hurdles

	Excluding latecomers (DV: Overall score)			Excluding incomplete evaluations (DV: Overall score)			Excluding unfinished evaluations (DV: Overall score)			Split sample analyses – Pre and post incentive (DV: Overall score)					
	(A42)*	**(A43)***	**(A44)***	**(A45)**	**(A46)**	**(A47)**	**(A48)**	**(A49)**	**(A50)**	**(A51)* Pre**	**(A52)* Post**	**(A53)* Pre**	**(A54)* Post**	**(A55)* Pre**	**(A56)* Post**
H1: Blind	-0.0802	-0.0149	-0.114	-0.0909	-0.0559	-0.131	-0.121	-0.0903	-0.158	-0.174	-0.00236	-0.179	0.0341	-0.130	-0.113
	(0.218)	(0.911)	(0.157)	(0.177)	(0.671)	(0.0947)	(0.0694)	(0.517)	(0.0404)	(0.0434)	(0.982)	(0.342)	(0.875)	(0.144)	(0.489)
Order		-0.00527			-0.00553			-0.00582				-0.00646	-0.00153		
		(0.205)			(0.159)			(0.163)				(0.283)	(0.787)		
HA1: Blind x Order		-0.00263			-0.00119			-0.000866				0.00135	-0.00259		
		(0.605)			(0.809)			(0.869)				(0.852)	(0.782)		
Female proposer			-0.426			-0.374			-0.320					0.671	-1.411
			(0.606)			(0.622)			(0.696)					(0.468)	(0.388)
Same subunit			0.118			0.177			0.153					0.153	0.235
			(0.497)			(0.275)			(0.358)					(0.440)	(0.447)
Same location			-0.238			-0.241			-0.253					-0.227	-0.336
			(0.267)			(0.185)			(0.208)					(0.507)	(0.231)
H2: Blind x Female proposer			0.0561			0.132			0.0954					-0.223	0.356
			(0.740)			(0.398)			(0.532)					(0.215)	(0.195)
H3: Blind x Same subunit			0.0112			0.0524			0.101					0.0132	-0.0720
			(0.965)			(0.808)			(0.679)					(0.972)	(0.855)
H4: Blind x Same location			0.260			0.124			0.127					0.0424	0.355
			(0.262)			(0.501)			(0.527)					(0.924)	(0.394)
Constant	2.847	2.501	2.851	2.681	2.475	2.696	2.652	2.450	2.676	2.143	2.943	2.099	3.075	2.104	3.005
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.0687)	(0.000)	(0.005)	(0.000)	(0.0542)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No	Yes	Yes
N	1617	1617	1617	1793	1793	1793	1728	1728	1728	997	840	997	840	997	840
R^2	0.4248	0.4103	0.4264	0.4024	0.3899	0.4054	0.4129	0.3985	0.4158	0.5852	0.6491	0.5506	0.6120	0.5874	0.6532

Notes. Dependent variables and sample restrictions are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models highlighted with an asterisk were not planned in the preregistration. Models A42 to A44 excluded evaluators that registered after the experiment had already started. Models A45 to A47 excluded all individual evaluations where evaluators did not rate an idea on all five outcomes. Models A48 to A50 excluded evaluators that did not evaluate all 48 ideas. Models A51 to A56 split the sample in two periods: pre and post the introduction of an extra incentive, which we announced roughly halfway through the experimental period. The incentive was twofold: We announced to donate 10$ to a charitable cause for each finished evaluation (of 48 ideas). In addition, we announced to randomly pick one innovation manager who had finished the evaluation to receive a gift card of $300.

TABLE A10 Online experiment: Participant descriptive statistics and balance checks

		Balance checks					
		Experimental designs			Between design only		
	All (1,473 participants)	**Within design** (737 participants)	**Between design** (736 participants)	**t-stat.**	**Non-blind condition** (368 participants)	**Blind condition** (368 participants)	**t-stat.**
Age	35.97	36.01	35.92	0.1781	35.46	36.38	-1.2237
	(9.96)	(9.75)	(10.17)	(0.8587)	(9.59)	(10.70)	(0.2215)
Female	0.49	0.48	0.50	-0.5464	0.47	0.52	-1.3268
	(0.50)	(0.50)	(0.50)	(0.5848)	(0.50)	(0.50)	(0.1850)
Education							
Tech or community college	0.13	0.12	0.14	-1.3403	0.15	0.13	1.0662
	(0.33)	(0.32)	(0.35)	(0.1804)	(0.36)	(0.33)	(0.2867)
Undergraduate	0.53	0.53	0.54	-0.7589	0.55	0.54	0.3696
	(0.50)	(0.50)	(0.50)	(0.4480)	(0.50)	(0.50)	(0.7118)
Graduate	0.29	0.31	0.26	2.2946	0.24	0.27	-1.0101
	(0.45)	(0.46)	(0.44)	(0.0219)	(0.43)	(0.45)	(0.3128)
Doctorate	0.05	0.05	0.06	-0.9365	0.05	0.06	-0.4709
	(0.22)	(0.21)	(0.23)	(0.3491)	(0.23)	(0.24)	(0.6378)
Occupational position							
Junior management	0.27	0.27	0.26	0.3378	0.26	0.26	-0.0837
	(0.44)	(0.44)	(0.44)	(0.7355)	(0.44)	(0.44)	(0.9333)
Middle management	0.46	0.47	0.45	0.9684	0.46	0.44	0.3700
	(0.50)	(0.50)	(0.50)	(0.3330)	(0.50)	(0.50)	(0.7115)
Upper management	0.17	0.17	0.18	-0.4923	0.17	0.18	-0.3838
	(0.38)	(0.38)	(0.38)	(0.6226)	(0.38)	(0.39)	(0.7012)
Researcher	0.10	0.09	0.11	-1.4209	0.11	0.11	0.1189
	(0.30)	(0.28)	(0.31)	(0.1556)	(0.31)	(0.31)	(0.9054)
Part-time	0.14	0.15	0.14	0.6611	0.13	0.14	-0.2149
	(0.35)	(0.36)	(0.34)	(0.5086)	(0.34)	(0.35)	(0.8299)
Working experience	14.98	15.11	14.86	0.4537	14.35	15.37	-1.3465
	(10.40)	(10.49)	(10.32)	(0.6501)	(9.72)	(10.88)	(0.1786)
US-based	0.35	0.33	0.36	-1.0044	0.36	0.35	0.3070
	(0.48)	(0.47)	(0.48)	(0.3154)	(0.48)	(0.48)	(0.7589)

Notes. Means and standard deviations (in parentheses) of the participants in the different versions and conditions of the experiment. The t-statistic (with p-value in parentheses) tests the equality of the means between the within and between designs and the non-blind and blind conditions (in the between design), respectively.

TABLE A11 Online experiment: Information about ideas

Idea title	Idea description	Name of Kickstarter project	Funding goal	Funding achieved	Number of backers	Date of funding
Learn a new language app	Learning a new language is nothing easy. Traditional language learning apps, like flashcards, are often based on sample sentences and images that may not speak to you. However, making digital or paper personalized notes would be effective but very time-consuming. So is there a way for people to learn language in a natural, fast and efficient way? This is an app that allows you to create personalized flashcards and mind maps to fasten your learning process. Research shows that by making choices about what information and in what format to store in your brain makes memorizing much easier. The whole design saves the redundant work of collecting word images and audios for you, but leaves the essential make-your-own-choice method. Moreover, it improves the flashcard function with the proven "spaced repetition method" which shows new and hard words more often to enhance memory	The Fluent Forever App: Learn to *Think* in Any Language	$250,000	$587,785	4,434	Oct 19 2017
Plan your gardening app	The pandemic and pollution definitely make "backyard revolution" a real plan. Now more people are planning to plant their own vegetables or crops in their backyard. However, new gardeners are often left alone once they received the seeds: When to seed? Bed preparing? Transplanting? Harvesting? Even experienced gardeners can get caught in a messy situation if there is a variety of species. While not everyone has a green-finger expert in the neighborhood to help out, how can we make it possible? This is a "personal gardening advisor" app that provides one-stop-shop services such as calendar planning, frost forecasting and gardening instructions. With an automatic weekly task list, even newbies can navigate through those tricky schedules, and does not worry about memorizing the complex schedules from seeding to harvesting. This app can even update your plans with local weather forecast. No more hustle, just follow your digital gardener advisor.	Seedtime: The Fastest Way to Plan Your Garden Ever	$30,000	$341,796	2,212	Apr 2 2021

App to allow you socialize digitally in a safe, private way	As a tradeoff to more social interaction, social media users give away their privacy. While the tech giants promise legal use, we still very often see scams and advertisements developed based on our private information. While some of us quit social media hoping for a happier and more focused life, this pandemic sets real barriers and shows us we still need social media to see our senior grandparents and little nephews. Then how can people socialize in the digital world in a safe, private way? This is a social media app that maintains the social function, but gives the users ultimate privacy protection. No strangers, no advertisement. But since this app does not make money out of selling your data, accordingly, a fee is justified to keep the service running. However, what makes it even easier is to collaborate with current social media platform and have a charged version so users do not need to transfer to the new platform.	Social Media is broken. Kohort is here to fix it.	$100,000	$110,673	323	Jan 16 2019
App for finding recipes	We all had similar frustration staring at unsorted ingredients in the fridge and having no idea what to cook for dinner. Cookbooks are not going to help, neither can food blogs. Both are too rigid for a last-minute home-cooking. Can I just have my granny over and turn these messy fridge leftovers into amazing meals through magic? This is a recipe generation app that gives you customized cook ideas by scanning the items in the fridge (well, typing is also okay). Using machine learning, the algorithm learns how to combine different ingredients and tastes and may even be able to generate your "signature dish" upon the unique combination of food in your fridge. Enjoy a joyful, sustainable and relaxing cooking session with your loved ones!	Kitchen Appetite: unlocking the magic of great meals	$75,000	$77,069	212	Mar 11 2019
Social traveling app	Travelling is not only about the new place, but also about the new people you meet on the way. However, there is no place in the local newspaper that allows the travelers to post: "Hey I am here from Monday to Wednesday, anyone wanna join me visiting the city?" So how can we bring travelers together and make a better travel experience for all? This is a traveler platform where users can create their own profile, share travel plans, and find travel pals. A unique feature is "wish list", where users can post activities and find people to join.	Bindle | The new way to travel	AU$ 12,000	AU$ 18,263	145	Sep 9 2014
App for preserving stories	Stories are lost when we forget them. Some are very precious, like the one from your grandmother or grandfather that you would like to hear again when they have passed away. Some are very encouraging, like the one of your friend battling with depression that might be helpful to a total stranger. But not everyone can go to a Ted Talk and throw a speech. Nor is your grandma or	StoryPress: Human History, Told through Personal Stories	$15,000	$15,299	176	May 1 2013

	grandpa comfortable with the pale "recorder" on the electronic device. So how can we preserve stories in a better way? This is a "story recording" app that leads the user through like a personal biographer. With a hearty setup with fireplace, users can pick the type of stories they would like to tell and then start with some leading questions to answer. All stories will be recorded in audio format. The app also allows users to upload their stories to the community so they can be shared, heard and saved.					
Dining out at discounted prices app	Rush-hour mismatching cost restaurants unearned revenue and diners bad experiences. During the downtime, restaurants are willing to offer some discounts and some diners do not mind the time, such as a happy-hour offer. Is it possible to match them together? This is a premium dining reservation app that posts restaurant discount offers, and allows diners to book and purchase ahead. Restaurants are able to tailor special menus without the hustle of changing the main menu in the shop, and diners can save money and have a nice experience	VIP Restaurant Booking APP	£20,000	£20,665	12	Dec 11 2020
App for finding people to do sports with	A great soccer game requires an evenly-competitive rival team and a nice sport arena. For amateur players who want to enjoy a sports game, it is crucial that they have access to local sports game information and is able to check the availability of sports arenas. This is a sports game platform that is specially designed for amateur players to join or organize a group sports game. The value lies in its simplicity and professionalism. Users are able to scroll down the sports games happening around their neighborhood, and register to join any of them. If users are an organizer, they are also able to check the local playground and book ahead.	Sport12 - Playing Amateur Sports	€7,499	€ 7.642	72	Mar 22 2019
App for starting a new lifestyle	There is as much motivation after reading a wellness/health blog post as hustle to download a habit-forming app and start it. But we all want a peaceful, inspirational life, so how can habit forming become simpler? This is an app that combines the content platform and habit-forming tools. The users can follow bloggers, read posts, and create their own to-do list or set up a running goal right now. This app saves all the hustles of finding the right app, creating a routine and daily check-in here. Simply choose the kind of goal you would like to reach-be it exercise, meditate, clean-eating, then this app will give you the nudge via a couple tricks (for example 21-day challenge, relative content feeding, community care etc.).	Thriver+: A positive habit forming app that simply works	AU$ 15,000	AU$ 17,096	179	Nov 6 2020

App for matching tutors and students	Finding a perfect tutor is hard. You can't just google it. Finding a suitable student is also hard, "I am not a big fan of you either" happens to the tutor too. While traditional ways rely on word-of-mouth, self-branding, or agency-recommendation, which all lack fairness and are of few choices, is there another way to get this matchmaking done? This is an online marketplace platform for tutoring. Building a market creates transparency, promotes efficiency and brings benefits to both parties. Tutors can register on the app, provide their experience and success stories, while students can "shop" around their favorite ones. And students can leave a review about the tutor, so future users can benefit from more objective judgements.	Nimbles: Find the best tutors and teachers!	€8,000	€ 10.034	120	May 20 2015

TABLE A12 Online experiment: Descriptive statistics

	Within design (7,332 evaluations)											
	Mean	**S.D.**	**Min**	**Max**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1 Blind	0.500	0.500	0	1								
2 Overall score	4.637	1.527	1	7	-0.02							
3 Next round	0.554	0.497	0	1	-0.01	0.69						
4 Desirability	4.730	1.660	1	7	-0.01	0.82	0.64					
5 Feasibility	4.877	1.483	1	7	-0.01	0.70	0.51	0.59				
6 Viability	4.608	1.558	1	7	-0.01	0.82	0.61	0.71	0.76			
7 Duration	54.148	58.439	8.051	1735.364	-0.02	0.02	0.04	0.02	0.00	-0.01		
8 Female name	0.251	0.434	0	1	-0.58	0.01	0.01	-0.00	0.01	0.00	-0.01	
9 Male name	0.249	0.432	0	1	-0.58	0.01	0.00	0.01	0.01	0.01	0.03	-0.33

	Between design (7,331 evaluations)											
	Mean	**S.D.**	**Min**	**Max**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1 Blind	0.500	0.500	0	1								
2 Overall score	4.581	1.551	1	7	0.00							
3 Next round	0.537	0.499	0	1	0.01	0.70						
4 Desirability	4.712	1.677	1	7	-0.01	0.81	0.65					
5 Feasibility	4.783	1.516	1	7	0.01	0.70	0.51	0.60				
6 Viability	4.523	1.588	1	7	-0.01	0.81	0.61	0.68	0.75			
7 Duration	55.675	62.536	8.017	1895.539	-0.04	-0.03	-0.02	-0.02	-0.03	-0.05		
8 Female name	0.249	0.433	0	1	-0.58	0.00	-0.01	0.00	-0.01	0.00	0.03	
9 Male name	0.251	0.433	0	1	-0.58	-0.00	0.01	0.01	0.00	0.01	0.02	-0.33

TABLE A13 Online experiment: Mean comparisons of treatment conditions

			Within design (7,332 evaluations)		
	Non-blind condition	**Blind condition**	**Diff.**	**t-stat.**	**p-value**
Overall score	4.666	4.608	0.0580	1.62	0.10
Next round	0.558	0.550	0.00796	0.68	0.49
Desirability	4.744	4.717	0.0267	0.69	0.49
Feasibility	4.892	4.862	0.0294	0.85	0.40
Viability	4.623	4.592	0.0309	0.85	0.40
			Between design (7,331 evaluations)		
	Non-blind condition	**Blind condition**	**Diff.**	**t-stat.**	**p-value**
Overall score	4.575	4.587	-0.0123	-0.34	0.74
Next round	0.533	0.541	-0.00807	-0.69	0.49
Desirability	4.730	4.693	0.0375	0.96	0.34
Feasibility	4.775	4.791	-0.0169	-0.48	0.63
Viability	4.535	4.511	0.0245	0.66	0.51

TABLE A14 Online experiment: Regression analyses

	Within design		Between design		Within design	Between design
	(A57)	**(A58)**	**(A59)**	**(A60)**	**(A61)**	**(A62)**
	Overall score	**Overall score**	**Overall score**	**Overall score**	**Overall score**	**Overall score**
H1: Blind	-0.0629		0.0124			
	(0.0497)		(0.717)			
H2: Female name		0.0696		-0.00273	0.0830	-0.0273
		(0.0877)		(0.948)	(0.175)	(0.656)
H2: Male name		0.0562		-0.0221	0.0863	-0.0614
		(0.129)		(0.603)	(0.151)	(0.325)
Female evaluator					0.286	0.214
					(0.000)	(0.000)
Female evaluator x Female name					-0.0734	0.0606
					(0.386)	(0.476)
Female evaluator x Male name					-0.0672	0.0964
					(0.419)	(0.260)
Constant	4.947	4.884	4.860	4.873	4.747	4.761
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	No	No	No	No
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N	7314	7314	7304	7304	7211	7124
R^2	0.1176	0.1176	0.1066	0.1066	0.0932	0.1136

Notes. Dependent variables are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level.

TABLE A15 Online experiment: Alternative dependent variables

	Within design								Between design							
	(A63) Next round	**(A64) Desirability**	**(A65) Feasibility**	**(A66) Viability**	**(A67) Next round**	**(A68) Desirability**	**(A69) Feasibility**	**(A70) Viability**	**(A71) Next round**	**(A72) Desirability**	**(A73) Feasibility**	**(A74) Viability**	**(A75) Next round**	**(A76) Desirability**	**(A77) Feasibility**	**(A78) Viability**
H1: Blind	-0.0152	-0.0349	-0.0430	-0.0355					0.00786	-0.0381	0.0170	-0.0239				
	(0.173)	(0.305)	(0.176)	(0.268)					(0.479)	(0.306)	(0.619)	(0.498)				
H2: Female name					0.0156	0.0251	0.0465	0.0432					-0.0152	0.0308	-0.0259	0.0193
					(0.263)	(0.557)	(0.238)	(0.275)					(0.262)	(0.502)	(0.535)	(0.657)
H2: Male name					0.0148	0.0448	0.0394	0.0278					-0.0005	0.0454	-0.00813	0.0286
					(0.270)	(0.271)	(0.291)	(0.464)					(0.969)	(0.320)	(0.847)	(0.509)
Constant	0.641	4.982	5.190	4.817	0.626	4.947	5.148	4.782	0.609	4.889	5.086	4.755	0.617	4.851	5.103	4.731
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
Idea FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7316	7329	7324	7285	7316	7329	7324	7285	7317	7330	7320	7283	7317	7330	7320	7283
R^2	0.0884	0.1050	0.0855	0.1185	0.0884	0.1050	0.0855	0.1185	0.0982	0.1007	0.0716	0.1005	0.0983	0.1007	0.0716	0.1005

Notes. Dependent variables are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level.

TABLE A16 Overview of key estimates from field and online experiment

	Field experiment	**Online experiment**	
		Within design	**Between design**
H1 (Blind)	-0.0989	-0.0629	0.0124
	(0.119)	(0.0497)	(0.717)
H2 (Female proposer)	0.109	0.0696	-0.00273
	(0.462)	(0.0877)	(0.948)
H3 (Unit)	0.0258	Not tested	Not tested
	(0.907)		
H4 (Location)	0.145	Not tested	Not tested
	(0.463)		

Notes. P-values in parentheses are based on standard errors clustered at the evaluator level.

TABLE A17 Online experiment: Alternative time thresholds

	Within design						Between design					
	8 seconds threshold (original)		5th percentile threshold		10th percentile threshold		8 seconds threshold (original)		5th percentile threshold		10th percentile threshold	
	(A79) Overall score	**(A80) Overall score**	**(A81)* Overall score**	**(A82)* Overall score**	**(A83)* Overall score**	**(A84)* Overall score**	**(A85) Overall score**	**(A86) Overall score**	**(A87)* Overall score**	**(A88)* Overall score**	**(A89)* Overall score**	**(A90)* Overall score**
H1: Blind	-0.0629		-0.0548		-0.0733		0.0124		-0.00861		-0.00447	
	(0.0497)		(0.0907)		(0.0281)		(0.717)		(0.805)		(0.901)	
H2: Female name		0.0696		0.0567		0.0788		-0.00273		0.0158		-0.000810
		(0.0877)		(0.169)		(0.0606)		(0.948)		(0.710)		(0.985)
H2: Male name		0.0562		0.0528		0.0678		-0.0221		0.00139		0.00976
		(0.129)		(0.160)		(0.0821)		(0.603)		(0.974)		(0.825)
Constant	4.947	4.884	4.916	4.861	4.932	4.858	4.860	4.873	4.836	4.828	4.846	4.841
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Evaluator fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
Idea fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Order fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	7,314	7,314	6,975	6,975	6,596	6,596	7,304	7,304	6,974	6,974	6,620	6,620
R^2	0.1176	0.1176	0.1239	0.1239	0.1301	0.1302	0.1066	0.1066	0.1088	0.1088	0.1114	0.1115

Notes. Dependent variables are given in the column headers. P-values in parentheses are based on standard errors clustered at the evaluator level. Models highlighted with an asterisk were not planned in the preregistration.

Appendix

Questionnaire for Paper 1.
The multi-national survey

Bringing AI to ops project

5 Markets – 1250 online interviews with decision makers/employees responsible for the digitization / automation of processes [Quota 1]

Market	Language	Age span	n =
China	Mandarin	18-65	250
Germany	German	18-65	250
India	English	18-65	250
UK	English	18-65	250
US	English	18-65	250

5 Markets – 1250 online interviews with operations managers/employees that implement and work with automated processes [Quota 2]

Market	Language	Age span	n =
China	Mandarin	18-65	250
Germany	German	18-65	250
India	English	18-65	250
UK	English	18-65	250
US	English	18-65	250

Quota setting and screening questions

Thank you for taking time to participate in this survey! Completion of this survey will take about 18 minutes.

S3. What best describes your current work situation?

1 ☐ Self-employed [CLOSE]
2 ☐ Full-time employed
3 ☐ Part-time employed (< 32 hours a week) [CLOSE]
4 ☐ Student [CLOSE]
5 ☐ Homemaker [CLOSE]
6 ☐ Unemployed [CLOSE]
7 ☐ Retired [CLOSE]
8 ☐ Other [CLOSE]

PROGNOTE: If working (S3 = 2). Single code

S4. What best describes the type of work that you do?

1 ☐ White collar worker (Management/Clerical/Administrative/Office worker, etc.)
2 ☐ Blue collar worker (skilled or semi-skilled Worker/Craftperson/Foreman, etc.) [CLOSE]
3 ☐ Other (unskilled worker/laborers, etc.) [CLOSE]

S5 – COMPANY SIZE INTERNATIONALLY SINGLE CODE	ASK ALL

How many people, including yourself, are permanently employed by your company **internationally**?

Please include all full and part time staff based at all sites of your company.

1	1-10 [CLOSE]
2	11-49 [CLOSE]
3	50-99 [CLOSE]
4	100-149
5	150-199
6	200-249
7	250-499
8	500-749
9	750-999
10	1000-1499
11	1500-1999
12	2000-2499
13	2500-4999
14	5000-9999
15	10000+

99	Not sure [CLOSE]	

S6. Does the company you work for produce, offer or sell products or services to other businesses or to consumers?

1. Yes, to other businesses
2. Yes, to consumers
3. Yes, to both
4. No [CLOSE]

****

We are interested in understanding the adoption and implementation of technologies that draw on data in order to generate new insights or to automate processes, such as AI, machine learning and advanced analytics. This includes sophisticated applications such as prediction models, data pattern recognition, digital assistants, image analysis software, speech and face recognition systems, just to mention a few.

For various reasons, such as improving efficiency or increasing productivity, more and more organisations strive to become more data oriented and introduce AI or advanced analytics in their work processes. Sometimes this goes quite easily and sometimes hurdles are encountered.

****

Ad1) Has your company implemented any kind of AI or advanced analytics tools?

- Yes, fully implement
- Yes, partly implemented
- Currently implementing
- No, not (yet) implemented [CLOSE]

Ad3) Which of the following best describes your work tasks?

1 ☐ I am part of a technical unit with responsibility to introduce AI or advanced analytics technologies into the company [Quota 1]

2 ☐ I am part of an operational unit that uses AI or advanced analytics in our operational processes. [Quota 2]

3 ☐ I don’t have any of the work tasks described above [CLOSE]

Main questionnaire

We are interested in your personal experience with AI or advanced analytics in your company. You do not need to have any technical knowledge about the tools to answer these questions.

PROGNOTE: Ask all.
Q1. When was the <u>most recent</u> initiative using AI or advanced analytics <u>anywhere</u> in your company initiated?
1. This year
2. 1-3 years ago.
3. 4-5 years ago.
4. 6 or more years ago. [CLOSE]
5. We have never done this. [CLOSE]
6. Don't know [CLOSE]

PROGNOTE: Ask all.
Q2a. When was the most <u>recent</u> initiative using AI or advanced analytics <u>in your unit</u>?
1. This year
2. 1-3 years ago.
3. 4-5 years ago.
4. 6 or more years ago.
5. We have never done this in my unit.
6. Don't know

PROGNOTE: Q2a 1/3.
Q3a. In how many initiatives in <u>your unit</u> have AI or advanced analytics been used during the last 5 years?
1. Less than 5.
2. 6-10.
3. 11-20.
4. 21-50.
5. 51-100.
6. More than 100.

Q2a 1/3.
Q3b. Out of these [insert from Q3a] initiatives <u>roughly</u> how many have faced challenges with implementation? Even if you're not sure, please give your best estimate

0% ------------ 100%

Q4a.
In how many initiatives <u>outside your unit</u> have AI or advanced analytics been used during the last 5 years?
1. Less than 5.
2. 6-10.
3. 11-20.
4. 21-50.
5. 51-100.
6. More than 100.
7. None that I am aware of

If Q4a > 1
Q4b. Out of these [insert from Q4a] initiatives <u>roughly</u> how many do you think have faced challenges of any kind? Even if you're not sure, please give your best estimate

0% ------------ 100%

PROGNOTE: Randomize order of Q5-Q7

PROGNOTE: Ask if Q3b or Q4b > 0. Randomize items, except last two items.
Q5. You indicated you have faced some challenges.

AI or advanced analytics challenges can sometimes be related to <u>technology</u>. Did your unit or company face any of the following challenges? (Select as many as apply.)

1. The tools are not adequate or too difficult to use
2. The technology is immature and unreliable
3. Procurement of the technology is too costly
4. Dedicated software/hardware is needed
5. The technology cannot produce relevant insights
6. This kind of technology is not suitable for our industry
7. This kind of technology could break laws and regulations governing our industry
8. We do not have access to the raw data needed to use this kind of technology
9. Data is not structured to enable use of AI or advanced analytics
10. Data definitions are not unified across the company
11. Other technology related issue: __________
12. None of the above [PROGNOTE: Can only be selected as single choice]

PROGNOTE: Ask if Q3b or Q4b > 0. Randomize items, except last two items.
Q6. You indicated you have faced some challenges.

AI or advanced analytics challenges can sometimes be related to <u>organizational structure</u>. Did your unit or company face any of the following challenges? (Select as many as apply.)

1. No clear ownership of AI or advanced analytics in the company
2. Budget or funding process lacking in this area
3. Short-term deliverables are prioritized over AI or advanced analytics
4. KPIs and reward structures are not set up to handle this kind of technology
5. There are too many silos in my company
6. The technical experts sit too far away from the operational domain
7. Lack of skilled employees for AI or advanced analytics related tasks
8. No clear ownership of data across my company
9. Other units don't have processes for managing data correctly
10. The company is too product focused
11. There are too many power struggles in our company
12. Lack of understanding among management
13. Lack of understanding among our customers
14. Uncertain market demand for customer-facing AI or advanced analytics solutions
15. Lack of external collaboration partners
16. Resistance to redesigning business processes
17. Other organization related issue: __________
18. None of the above [PROGNOTE: Can only be selected as single choice]

PROGNOTE: Ask if Q3b or Q4b > 0. Randomize items, except last two items.
Q7. You indicated you have faced some challenges.

AI or advanced analytics challenges can sometimes be related to <u>people and culture</u>. Did your unit or company face any of the following challenges? (Select as many as apply.)

1. Many employees prefer sticking to tried and tested routines
2. Many employees are not open to change in general

3. Employees do not see what value AI or advanced analytics bring to the business
4. Employees do not see how AI or advanced analytics improve their job satisfaction
5. Employees are afraid of losing their jobs if these technologies take over
6. Employees worry that their expertise will be ignored or made useless
7. Employees worry that they will be less in control
8. Employees don't trust the output from AI or advanced analytics
9. The technical and operational teams do not understand each other
10. Older generations don't understand the technology
11. Our company culture is not open to change
12. Unions are against the adoption of these technologies
13. Other people or culture related issue: __________
14. None of the above [PROGNOTE: Can only be selected as single choice]

PROGNOTE: Randomize order of Q8-Q10

PROGNOTE: Ask if "None of the above" is not selected in Q5. Randomize items, except last two items.
Q8. You indicated you have faced some technical challenges.

How has your company tried to overcome the <u>technology</u> challenges you have faced with AI or advanced analytics? (Select as many as apply.)

1. Improve usability of tools
2. Use technology that is less cutting edge
3. Find cheaper solutions
4. Search for more standardized solutions
5. Try to improve relevance and quality of output
6. Find tools and technologies specific to our industry
7. Find tools with higher transparency or compliance with regulations
8. Improve access to data sources
9. Setting up a data lake or similar for centralized data access
10. Setting up data management policies to ensure proper handling of data within the company
11. Other way to overcome technology challenges: __________
12. We have not tried to overcome technology challenges [PROGNOTE: Can only be selected as single choice]
13. Don't know

PROGNOTE: Ask if "None of the above" is not selected in Q6. Randomize items, except last two items.
Q9. You indicated you have faced some <u>organizational</u> challenges.

How has your company tried to overcome <u>organizational</u> challenges you have faced with AI or advanced analytics? (Select as many as apply.)

1. Define new organizational units to clarify ownership of AI or advanced analytics
2. Implement clear budget or funding processes
3. Set up an AI or advanced analytics process parallel to current practice to compare
4. Set up clear evaluation criteria for new AI or advanced analytics projects
5. Set up KPIs and score cards that quantify gains from AI or advanced analytics
6. Build a best practice team using external hires with experience from other industries
7. Build cross functional teams
8. Create work roles focused on increasing collaboration across organizational units
9. Promote a data driven workflow across organizational boundaries
10. Create management supported evangelist roles for AI or advanced analytics

11. Promote understanding of AI or advanced analytics among your customers
12. Develop new business areas where AI or advanced analytics is used from the very start
13. Other way to overcome organizational challenges: __________
14. We have not tried to overcome organizational challenges [PROGNOTE: Can only be selected as single choice]
15. Don't know

PROGNOTE: Ask if "None of the above" is **NOT** selected in Q7. Randomize items, except last two items.
Q10. You indicated you have faced some <u>people or culture</u> challenges.

How has your company tried to overcome <u>people or culture</u> challenges you have faced with AI or advanced analytics? (Select as many as apply.)

1. Educate employees on benefits of work routine change
2. Hire younger employees or employees that have experience from fast changing work environments
3. Set up technology workshops, training sessions or educational courses for employees
4. Present AI or advanced analytics using proofs of concept and demos
5. Making executive managers openly share their support for the technology
6. Use "shadow" projects where employees try out AI or advanced analytics and compare with current work practices
7. Tie career development goals to the use of AI or advanced analytics
8. Find career paths for employees who embrace fast workplace change
9. Encourage employees to develop new skills tied to AI or advanced analytics
10. Work with unions to see employee benefits from AI or advanced analytics
11. Other way to overcome people or culture challenges: __________
12. We have not tried to overcome people of culture challenges [PROGNOTE: Can only be selected as single choice]
13. Don't know

PROGNOTE: Ask if more than one selection was made across all of Q5 – Q7.
Q11. You mentioned the following challenges faced by your unit or company with AI or advanced analytics. Overall, which are the most critical challenges? (Select top three challenges.)

PROGNOTE: List all challenges selected across all of Q5 – Q7.

PROGNOTE: Ask if more than one selection was made across all of Q8 – Q10.
Q12. You mentioned the following ways you have tried to overcome challenges with AI or advanced analytics in your unit or company. Overall, which are the most effective strategies to overcome the challenges you have faced? (Select top three strategies.)
PROGNOTE: List all strategies selected across all of Q8 – Q10.

Q13.
What are the main reasons for adopting AI or advanced analytics into your company? (Select all that apply.)
[PROGNOTE: Randomize items]

1. Profitability
2. Productivity
3. Revenue increase
4. Cost reduction

5. Workforce reduction
6. More innovative products and services
7. Improving customer experience
8. Other: ______
9. Don’t know

Q14.
How does AI or advanced analytics influence your company going forward; To what extent do you agree with the following statements?
[7 grade scale, from completely disagree to absolutely agree]
[PROGNOTE: Randomize items]

1. AI or advanced analytics will make our current processes become obsolete
2. Becoming AI-driven, we will need to form new KPI's to measure our performance
3. Employees will still have the same roles as they have today
4. Most employees in the company will use AI or advanced analytics
5. Managers will still have the same roles as they have today
6. All employees will be expected to come up with new ideas for AI or advanced analytics
7. Using AI or advanced analytics will not change our company significantly
8. There will be a constant flow of new AI or advanced analytics applications in the company
9. Working with AI or advanced analytics will lead to a continuous redesign of work processes
10. Over time more and more work tasks within our company will be overtaken by AI or advanced analytics
11. In the long term our focus will shift from producing products and services to producing AI algorithms and models
12. There are only a limited number of processes within our company that will benefit from AI or advanced analytics
13. AI or advanced analytics will result in more frequent changes in the organizational design/setup of my company
14. We have a clear vision of how the company will be structured when we have finished implementing AI or advanced analytics
15. Using AI or advanced analytics will make job tasks more complex
16. AI or advanced analytics will allow us to mainly employ low wage workers
17. AI or advanced analytics will allow us to mainly employ experts

Q15. When do you estimate your company will have completed the organizational transformation needed to become fully AI or advanced analytics driven?

1. This year
2. In ____ years (enter number of years)
3. Never, this will be a process of constant change
4. Never, we are not aiming to be fully AI or advanced analytics driven

Q16a. Which of the following roles do you have employed within your company? (Select as many as apply.)

1. Data Scientist
2. Data Engineer
3. Data Analyst
4. Data Architect
5. Machine Learning Engineer
6. Data Manager
7. Business Translator (link between technical and business units)
8. AI Evangelist/Ambassador
9. Other related role, please specify:
10. None of the above

11. Don't know

Q16b. Which of the following roles do you plan to have within your company the coming 3 years? This will be in addition to the roles you currently have in place. (Select as many as apply.)

1. Data Scientist
2. Data Engineer
3. Data Analyst
4. Data Architect
5. Machine Learning Engineer
6. Data Manager/Steward
7. Business Translator (link between technical and business units)
8. AI Evangelist/Ambassador
9. Other related role, please specify
10. No other related role
11. Don't know

Q17.
What is the level of maturity of AI or advanced analytics in your company? To what extent do you agree with the following statements?

[7 grade scale, from strongly disagree to strongly agree]
[PROGNOTE: Randomize items]

a. We use proof of concept or pilot projects extensively to test AI or advanced analytics
b. AI or advanced analytics projects are put into production and used as best practice on a regular basis
c. AI or advanced analytics projects always have an executive sponsor and a dedicated budget
d. AI or advanced analytics is considered for use in all new projects, products and services
e. We have a clear strategy for AI or advanced analytics related data management.
f. AI or advanced analytics are implemented throughout the company

Q18. How important are the following sources for your AI or advanced analytics activities?
[7 grade scale, from very low importance to very high importance]
[PROGNOTE: Randomize items]

a. Suppliers of equipment, materials, components
b. Suppliers of software
c. Suppliers of data
d. Clients or customers
e. Competitors
f. Consultants
g. Commercial laboratories/R&D enterprises
h. Universities or other higher education institutes
i. Government research organizations
j. Other public sector
k. Private research institutes
l. Enterprises within your enterprise group
m. Professional conferences, meetings
n. Trade associations

o. Technical/trade press, computer databases
p. Fairs, exhibitions
q. Technical standards

Q19. The following questions deal with the use of data. To what extent do you agree with the following statements?
[7 grade scale, from completely disagree to absolutely agree]
[PROGNOTE: Randomize items]

a. My company uses data published by other companies in order to improve our own AI or advanced analytics activities
b. My company shares data so that other companies can use it in their business activities
c. My company freely shares some data with no restrictions or associated costs.
d. My company has a good understanding of licenses governing the use of data

Q20. How important to the management of your business are the following methods of organizing AI or analytics driven work?
[7 grade scale, from very low importance to very high importance]
[PROGNOTE: Randomize items]

a. Planned job rotation of staff across different functional areas
b. Regular brainstorming sessions for staff to think about improvements that could be made within the business
c. Cross-functional work groups or teams (combined across different working areas or functions)

Q21. In my unit, over the next 18 months, we plan to invest in AI or advanced analytics …
[7 grade scale, from 'significantly less than previously' to 'significantly more than previously']

Q22. In my unit, over the next 18 months, we will directly invest in AI or advanced analytics with…
[7 grade scale, from 'a significantly narrower focus' to 'a significantly wider scope']

Demographics & Classification

Your responses will be completely confidential and anonymous and will only be combined with the responses of other people.

S1. Please specify your gender.
1 ☐ Male
2 ☐ Female
Other
Refuse to say

PROGNOTE: Ask all. Single code; valid: 18-65
S2. How old are you? _ _
refused

PROGNOTE: Ask all. Single code. ***use local list according to school system***
Q38. What is the highest level of education you have completed?

1.	No school
2.	Primary school
3.	Junior/Middle school

4.	High school
5.	Trade studies
6.	College/University
7.	Post graduate degree

Q39 – INDUSTRY SECTOR SINGLE CODE	ASK ALL

Which of these industry sectors best describes the main activity of your company?

1	Agriculture, forestry, fishery	
2	Automotive	
3	Chemical	
4	Construction and engineering	
5	Education	
6	Financial services / Banking	
7	Government (Federal, State or Local)	
8	Healthcare	
9	Insurance	
10	IT / technology products	
11	Journalism	
12	Leisure/tourism/hospitality (restaurant/hotel)	
13	Manufacturing or processing industry	
14	Mining, oil, gas & resource extraction	
15	Not for profit / charities	
16	Pharmaceuticals/Biotechnology	
17	Professional services	
18	Real-Estate	
19	Retail and E-commerce	
20	Telecoms	
21	Transport, logistics & distribution	
22	Utilities	
23	Wholesale	
24	Water & waste management	
98	Other (specify)	
99	Don't know	CLOSE

S7. Which of the following best describes your role in your company?

1. I have a management role, with subordinates that report to me
2. Hierarchically I have people working under me, but I am not formally responsible for their function (e.g. set their salary).
3. I don't have people working under me.

T4b – TYPE OF DECISION MAKER SINGLE CODE	IF S7=1

What best describes your role at your place of work?

1	Senior management level – Being the one with the final say-so in strategic and operational decisions	Decision Maker
2	Management level – Significant influence, refers to making a recommendation or being consulted directly in strategic and operational decisions	Decision Maker
3	Operationally involved management – Some influence, refers to being one of many who make a suggestion with only moderate possibility of the suggestion being adopted.	Decision Maker