BerkeleyHaas
Haas School of Business
University of California Berkeley

Organizational Decision-Making Structures in the Age of Artificial Intelligence

California Management Review
1–18

DOI: 10.1177/0008125619862257
journals.sagepub.com/home/cmr
SAGE

Yash Raj Shrestha[1], Shiko M. Ben-Menahem[1], and Georg von Krogh[1]

SUMMARY

How does organizational decision-making change with the advent of artificial intelligence (AI)-based decision-making algorithms? This article identifies the idiosyncrasies of human and AI-based decision making along five key contingency factors: specificity of the decision search space, interpretability of the decision-making process and outcome, size of the alternative set, decision-making speed, and replicability. Based on a comparison of human and AI-based decision making along these dimensions, the article builds a novel framework outlining how both modes of decision making may be combined to optimally benefit the quality of organizational decision making. The framework presents three structural categories in which decisions of organizational members can be combined with AI-based decisions: full human to AI delegation; hybrid—human-to-AI and AI-to-human—sequential decision making; and aggregated human–AI decision making.

KEYWORDS: decision making, artificial intelligence, algorithms, organizational structure, delegation

How to structure organizational decision making—that is, designing where, when, and how to make and integrate decisions involving groups of individuals[1]—has long been a cornerstone concern in organization theory and micro-economics.[2] Herbert Simon defined rational decision making as the process of selecting the alternative that is expected to result in the most preferred outcome.[3] This process involves identifying and listing the alternatives, estimating their consequences, and comparing the accuracy and efficiency of each of these consequences. Organizations can be viewed as "networks of decisions"[4] that need to be structured in such a way as to

[1]ETH Zürich, Zürich, Switzerland

best attain organizational goals. Choosing the most appropriate decision-making structure—for example, delegating decisions to experts or aggregating the decisions of a group of individuals—has important implications for organizational performance.

While the challenges of designing decision-making structures involving human actors are fairly well understood, the recent rise of decision making by artificial intelligence (AI) algorithms introduces a new set of challenges to this age-old problem.[5] By synthesizing robust patterns from large data sets, AI—and, in particular, machine learning algorithms—enables the creation of new information and predictions from data (provided that the future can be fairly well predicted by existing data). The promise of fast, accurate, repeatable, and low-cost decisions, with quality approaching human-like intelligence, has been an important driver of the rapid developments in AI.[6] Indeed, experts in various professions—including medicine (e.g., surgery allocation), psychological counseling (e.g., therapeutic conversational agent), human resource management (e.g., hiring decisions), banking (e.g., credit risk predictions), science (e.g., astronomy), transportation (e.g., self-driving vehicles), public administration (e.g., immigration decisions), and legal counseling (e.g., bail decisions)—increasingly rely on the guidance of AI-based algorithms when making important decisions.[7]

While the rapid adoption of AI attests to the many measurable benefits of AI's learning and prediction power, its application in organizational decision making needs to be based on an adequate understanding of its strengths and weaknesses.[8] Indeed, managers who involve AI in decision making ultimately remain responsible for decision outcomes. Yet, recent events and mounting evidence from research show that the application of AI-based decision making may introduce and amplify a host of grave and often hidden biases and challenges for upholding fairness, accountability, transparency, and, consequently, trust in AI-based decisions.[9] Thus, although the appeal of AI-augmented human decisions has raised high expectations, how to design organizational structures that combine human and AI-based decision making so as to maximize its benefits and minimize risks remains an open question.[10]

In this article, we address this lacuna in the literature by building a framework that addresses the practically relevant question: *What is the most appropriate organizational structure for decision making involving AI?*

How Do Human and AI-Based Decision Making Compare?

Before addressing the organizational structures through which human and AI-based decision making can be combined, we compare their characteristics along five key decision-making conditions: specificity of the search space, interpretability of the decision-making process and outcome, size of the alternative set, decision-making speed, and replicability. Table 1 summarizes the characteristics of human and AI-based decision making along these conditions.

TABLE 1. Comparison of AI-Based and Human Decision Making.

Decision-Making Conditions	AI-Based Decision Making	Human Decision Making
Specificity of the decision search space	Requires a well-specified decision search space with specific objective functions.	Accommodates a loosely defined decision search space.
Interpretability of the decision-making process and outcome	Complexity of the functional forms can make it difficult to interpret the decision process and outcomes.	Decisions are explainable and interpretable, though vulnerable to retrospective sense-making.
Size of the alternative set	Accommodates large alternative sets.	Limited capacity to uniformly evaluate a large alternative set.
Decision-making speed	Comparatively fast. Limited trade-off between speed and accuracy.	Comparatively slow. High trade-off between speed and accuracy.
Replicability of outcomes	Decision-making process and outcomes are highly replicable due to standard computational procedure.	Replicability is vulnerable to inter- and intra-individual factors such as differences in experience, attention, context, and emotional state of the decision maker.

Specificity of the Decision Search Space

Because AI algorithms make decisions based on computational optimization, the "space" wherein the decision is searched needs to be carefully specified and restricted in terms of the objective function. Consider, for example, an AI algorithm designed to propose to human decision makers the "best" candidate from a set of applicants. This process demands a specific definition of the desired qualities and characteristics that need to be optimized—such as a candidate's predicted long-term productivity after hiring and sociability with other team members—as well as a set of variables that should be considered for selection, such as the candidate's education level, age, and domain of expertise. Today's AI technology is limited to well-structured (modular or "decomposable") decision objectives and is thus often referred to as "narrow" or "weak" AI. While Artificial General Intelligence—a "strong" version of AI capable of performing any type of decision—has drawn substantial research attention in recent years, experts agree that this technology will take several more years to mature and achieve the desired level of accuracy.[11] Human decision makers, in contrast, can exercise judgment and intuition in decision making and can thus address ill-structured decision objectives—often with counter-intuitive decision decompositions. As a result, decision making by humans may be difficult to explicitly describe (code) by an objective function.[12] In the hiring example, a human decision maker may "intuitively" base their decision on a set of tacitly held preferences (e.g., fit of the candidate with the organizational culture) without being able to explain why and with what weights such criteria were considered.[13]

Interpretability

Current AI algorithms typically identify patterns in data using automated search processes that result in an optimal prediction model. This search-for-patterns process usually involves so-called local optimization techniques (e.g., stochastic gradient descent)[14] where an objective function is incrementally optimized at each step of the algorithm. However, such procedures do not provide a holistic explanation of how AI arrives at its decision. Moreover, as the entire procedure is automated, identified patterns and models can have extraordinary complexity.

In our recruiting example, a well-performing model could have learned that the value of the variable education to the power of five interacted with the candidate's age to the power of nine is an important predictor of sociability. Because such a pattern would have emerged without any explanation and is thus difficult to interpret, AI algorithms are often referred to as "black box" models. The lack of interpretability of AI-based decision-making algorithms makes it difficult to identify biases embedded in the algorithmic process, and consequently, generate trust in AI-based decision outcomes.[15] This is particularly problematic in applications of deep learning algorithms, which typically combine the behavior of single nodes in hundreds of layers of neural networks. The opacity of algorithms also leaves AI-based decisions vulnerable to concealed tampering and adversarial attacks.[16]

Human decision makers can more readily backtrack their reasoning steps and provide explanations and justifications for why they made a certain decision. Yet, while explanations or narratives of decision-making processes may be more comprehensible, they may not always be accurate, truthful, or comprehensive.[17] For example, when asked why a certain job candidate was selected, human decision makers may find it difficult to disentangle the set of factors they considered. Indeed, there exists robust evidence that human decision makers are prone to provide distorted retrospective accounts of situations and decisions and hold biases that are relatively inaccessible to others.[18]

Alternative Set Size

Because AI-based algorithms use an automated search for the best fitting model, they can be used to evaluate the same set of objective functions uniformly and consistently over millions of alternatives. For example, once it is defined what constitutes the "best" candidate, the same criteria can be autonomously and efficiently evaluated over millions of applicants. Human decision making is limited by cognitive constraints that make it practically impossible to uniformly process large numbers of alternatives. When a large number of seemingly equivalent alternatives are available, human decision makers quickly become overwhelmed with the multitude of potential outcomes and the inherent risks that may result from making the wrong choice ("choice overload").[19] A larger alternative set increases the likelihood that the decision maker will make the wrong choice—leading to cognitive dissonance,

a state of mental discomfort where the decision maker holds multiple contradictory beliefs.[20] An overload of alternatives might also result in an inability to decide ("paralysis by analysis").[21]

Decision-Making Speed

Advances in computing hardware—particularly in general processing units and tensor processing units—and efficient algorithms have enabled AI-based decision making to occur at a near-instantaneous speed.[22] This algorithmic feature has made great impact on decision making in high-velocity contexts, such as high-frequency foreign exchange trading. The need to make speedy decisions can be detrimental to human decision-making outcomes.[23] Under high time pressure, decision makers often utilize heuristics to overemphasize some and ignore other information, leading to a speed-accuracy trade-off.[24] Indeed, Kahneman distinguishes human decision making into System 1 thinking—fast, intuitive, automatic, unconscious, and effortless—and System 2 thinking—slow and deliberate. System 1 makes decisions quickly by relying on heuristics such as associative thinking. Therefore, decision making in high-speed environments that activates System 1 can be subject to substantial deviations from reality and be vulnerable to systematic errors.[25] Researcher have also discovered an inverse relationship between the amount of time it takes to deliberate on a decision and a decision maker's confidence in that decision.[26]

Replicability

AI algorithms follow standard and non-ambiguous—yet relatively inflexible—decision processes that provide consistent outcomes given consistent inputs.[27] Human decision making, in contrast, involves inter- and intra-individual variance in experience, attention patterns, emotions, and information processing that influence the type of information individuals attend to, encode, and retrieve when making decisions. Such idiosyncrasies make replication of results highly problematic.[28] For example, psychology research has shown that decision fatigue may lead to deteriorating quality of decisions as an individual's mental energy is gradually depleted,[29] and research in cognitive science and neuroscience has shown that emotions constitute powerful and sometimes unpredictable factors in decision making.[30]

Combining Human and AI-Based Decision Making: Three Decision-Making Structures

Based on our comparison of human and AI-based decision making, we provide a framework that outlines how both modes may be combined to optimally benefit the quality of organizational decision making. Our framework (see Table 2) comprises three structural categories: full human to AI delegation; hybrid—human-to-AI and AI-to-human—sequential decision making; and aggregated human–AI decision making.

TABLE 2. Organizational Decision-Making Structures Involving AI-Based Algorithms.

Organizational Structure	Specificity of the Decision Search Space	Interpretability	Size of the Alternative Set	Decision-Making Speed	Replicability	Examples
Full human to AI delegation	High (required for AI to function)	Low (due to absence of human involvement)	Large (not restricted by human capacity)	Fast (not restricted by human capacity)	High (computationally standardized)	Recommender systems, digital advertising, online fraud detection, dynamic pricing.
Hybrid 1: AI to human sequential decision making	High → Low (high in the first phase, low in the second phase)	High (due to human involvement in the final decision)	Large (due to involvement of AI in the first phase)	Slow (due to human decision-making as a bottleneck)	Low (vulnerable to human variability)	Idea evaluation, hiring.
Hybrid 2: Human to AI sequential decision making	Low → High (low in the first phase due to human involvement, and high in the second phase for AI)	Low (due to AI involvement in the final decision)	Small (due to human involvement in the first phase)	Slow (due to human decision-making as a bottleneck)	Low (vulnerable to human variability)	Sports analytics, health monitoring.
Aggregated human–AI decision making	Low (for decisions allocated to humans) High (for decisions allocated to AI)	High (for decisions allocated to AI) Low (for decisions allocated to humans)	Small (same set of alternatives are evaluated by both humans and AI)	Slow (due to human decision-making as a bottleneck)	Partial (replicability only guaranteed in decision elements allocated to AI)	Top management teams, boards.

Full Human to AI Delegation

In designs involving full delegation of decision making, AI-based algorithms make decisions without human intervention—similar to organizational settings where managers delegate decision-making authority to human experts. Human decision makers, however, still retain responsibility for the decision. Full delegation is particularly useful in decision-making scenarios where the decision search space is specific and restricted, interpretability of the decision-making process is less important than the accuracy of the prediction, the alternative set is large, decision-making speed is critical, and replicability of decision outcomes is desirable.

While pure forms are still limited, current applications of full delegation to AI include traffic planning, real-time product recommender systems, dynamic pricing (e.g., pricing in airlines and hotels, high-frequency trading), and online fraud detection. In all of these examples, algorithm designers can accurately specify a concrete objective function. For example, recommender systems—such as those used to recommend products (e.g., Amazon) or streaming video (e.g., YouTube and Netflix)—are designed to maximize consumer engagement, sales, and ad revenues, while fraud detection systems are designed to detect unexpected activity and minimize losses. To perform these objectives, AI-based algorithms instantaneously scan and evaluate millions of data points for millions of users—a process that would be practically impossible with human involvement. Stability in the data generation process and the possibility to specify and restrict the decision search space is necessary for the AI-based decision-making algorithms to perform accurately. As the aggregate patterns of the behavior of many users do not change radically over time (while that of some users can), AI-based systems are able to predict preferences of classes of users at scale with high accuracy.

The premium placed on decision-making speed and optimization of the objective function typically involves a trade-off with human interpretability. Recommender systems, for example, can be designed to improve themselves without a human designer's understanding of the mechanism underlying the improvement. Using large amounts of data on granular user interactions and instantaneous feedback from user of digital platforms enable AI algorithms to learn user behavior such that decision-making efficiency and accuracy increase over time. Recommender systems can improve their performance (e.g., user engagement, profit maximization) by allowing machine-learning algorithms to automatically identify a set of features (e.g., placement order or suggestions) that influences the algorithm's performance. The algorithm then experiments by tuning (i.e., increasing or decreasing) those features and observing their influence on performance—a process that involves techniques such as randomized confirmatory tests and user experiments.[31] Similar conditions apply to real-time dynamic pricing in the airline and hotel industry, ad auctions, and high-frequency trading—where the speed of bidding and/or buying is critical. In all of these settings, human involvement would induce a debilitating delay in decision making and, most likely, reduce decision-making quality.

Full human to AI delegated decision making also involves several critical limitations. Studies have shown that machine-learning algorithms can acquire and replicate implicit human biases toward race and gender from the online textual data they use to derive insights and inform their decisions. For example, scholars have documented how popular online translation systems—which perform natural language processing using statistical machine translation (SMT)—construct gender-stereotyped translations from gender-neutral languages. To illustrate, Caliskan and colleagues note that Google Translate translates the Turkish gender-neutral "O bir doktor. O bir hemşire." to these English sentences: "He is a doctor. She is a nurse."[32] In a similar way, search engine results, Google's autocomplete function, and Facebook ads have been shown to embed hateful query suggestions and negative biases against women of color, religious groups, and the poor.[33] These examples show that, left unchecked, AI-based decision making may not only perpetuate but amplify cultural stereotypes and discrimination.[34]

In addition to these concerns, organizational structures with fully delegated decision making may come under scrutiny due to the design ethics of managers and computer engineers. Indeed, minimizing harmful outcomes of automated decisions require ethical choices in the objectives of the algorithms, data collection methods, data cleaning and pre-processing, feature selection, simulation of algorithm behavior, and data representation. As recent examples show, algorithms designed with the narrow objective of maximizing user engagement and ad revenues can expose users—and society at large—to dangerous vulnerabilities and harmful consequences for public well-being and democracy. As a case in point, YouTube's recommender system has come under fire for steering users to misleading material and inflammatory videos. Such content—while optimizing viewers' attention, engagement, and consequently, the company's ad revenues—have been linked to radicalization of viewers and divisiveness.[35] Similarly, research has shown that the personalization algorithms that curate and filter newsfeeds on social media platforms such as Facebook and Twitter and news aggregators such as Google News have contributed to a dynamic in which users are increasingly exposed to less diverse points of view.[36] The creation of such "filter bubbles" or "echo chambers" has been argued to foster perilous polarization and the spread of misinformation in society.[37]

Addressing these concerns requires the joint efforts of policy makers, the academic community, business leaders, and designers of algorithmic decision-making systems. Such efforts begin with the realization that managers can delegate authority to AI, but not responsibility. Thus, to reap the benefits while minimizing the risks of full delegation to AI-based decision makers, business leaders should both develop an understanding of how emerging legal frameworks such as the European General Data Protection Regulation (GDPR) may affect algorithmic quality, fairness, accountability, and transparency and take a proactive stance in ensuring the ethical design of algorithmic decision making. Such efforts include adopting novel solutions for debiasing data and contributing to the development of new methods for fair, accountable, and transparent algorithms.[38]

Hybrid Sequential Decision-Making Structures

Hybrid decision-making structures concern organizational designs where humans and AI-based algorithms sequentially make decisions such that the output of one decision maker provides the input to the other.[39] Hybrid structures enable organizational designers to benefit from the strengths of both human and AI-based decision making, yet may also amplify each other's weaknesses. We next consider two stylized hybrid structures: algorithmic decisions as input to human decision making and human decisions as input to algorithmic decision making.

Algorithmic decisions as input to human decision making. This structure consists of two phases. In the first phase, AI-based decision making is applied to the initial set of alternatives. AI functions as a filter that rejects redundant or inappropriate alternatives and passes a subset of suitable alternatives to the second phase in which a human decision maker selects from these alternatives. Placing AI-based decision making in the first phase allows human decision makers to effectively handle situations involving a large set of alternatives. This structure is analogous to the process whereby expert advisors offer recommendations on a set of alternatives to a decision maker with authority over the final decision, allowing decision makers to exercise discretion with respect to whether or not they take an expert's advice into consideration.[40] Similar to the full delegation design, effective functioning of AI in this structure requires specificity in the decision search space. While human involvement renders the decision more interpretable, the decision-making process loses replicability and speed.

This structure finds applications in crowd sourcing contests, healthcare monitoring, hiring, and loan application assessment. Crowdsourced innovation contests, for example, enable firms to involve large groups of individuals from outside the firm in the search for solutions to its problems.[41] By formulating a problem and broadcasting it to the crowd, firms can attract a diverse set of solutions. In so doing, the cost of problem-solving shifts from generating solutions to evaluating and selecting solutions. Sifting through a large set of solutions is tedious, time consuming, and costly. Using AI to categorize solutions, differentiate among various alternatives, and suggest a narrower alternative set allows human decision makers to evaluate solutions more efficiently. Moreover, for each decision, the algorithm can be configured to calculate and inform the confidence level of its suggestions.

Similar to the full delegation structure, designs in which human decision makers rely on the inputs of AI-based decision-making algorithms are vulnerable to certain errors and biases. Importantly, AI-based decisions involve the risk of omission errors in which viable alternatives are discarded (false negatives). Because rejections are automated, discarded alternatives remain concealed from human decision makers. Moreover, given that AI-based selection decisions are trained on prior human decisions, rejections are prone to reproducing institutional and systemic biases that subsequently feed into human decisions. For example, an AI recruiting tool developed by Amazon to identify promising job

candidates was found to output decisions biased against women. Trained on résumés the company received over a 10-year period, the computer model learned to favor male candidates on the basis of the overrepresentation of male candidates in technical roles during that period. As a result, Amazon's AI taught itself to penalize résumés from candidates from all-women's colleges and other gender identifying content. Following internal and external condemnation, Amazon discontinued the project.[42] In another alarming incident, Angwin and colleagues found that COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)—an algorithmic system used in U.S. courts to estimate the risk of recidivism and support bail and sentencing decisions—was biased against black defendants. As the tool's error rates were asymmetric, black people were more vulnerable to be incorrectly labeled as higher-risk compared with white defendants.[43] Such examples call for caution with blind confidence in AI-based recommendations to human decision making in settings where the right for equal treatment and equal opportunities is at risk.[44]

These limitations notwithstanding, researchers in machine learning have been actively working on developing AI systems that learn responsibly by making decisions only when its predictions are reliably aligned with the system's objectives, considering both accuracy and fairness. Such algorithms enable a more reliable collaboration with human decision makers. For example, research has shown that designing an algorithm to learn to defer or choose to pass a decision on to a human agent can greatly improve the accuracy and fairness of an entire system. Researchers have also designed AI systems that when working in tandem with a human decision maker can adaptively learn to defer decision making depending on both its confidence in the model's accuracy and the human decision maker's expertise and weaknesses.[45]

Human decisions as input to algorithmic decision making. In this structure, human decision makers first select a relatively small set of alternatives from a larger pool of alternatives, and then pass this set on to AI algorithms for evaluation and selection of the best alternative. This structure is effective in scenarios where human decision makers have high confidence in a small set of preferred alternatives, but the effective evaluation of this small set either requires the processing of large amounts of data and careful attention of decision makers over long period of time. This structure can effectively exploit the predictive capability of algorithms in situations where humans are uncertain about the best alternative out of the selected small set of alternatives. Because this structure relies on human decision making in the first phase, it is suitable for settings in which the size of the alternative set is small. AI-based decision making in the second phase requires high specificity of the decision search space. The optional involvement of human decision making as the third step allows for the final decision to be interpretable, yet as in the case of AI-to-human structure, this step reduces decision-making speed and replicability.

Billy Beane, the manager of the professional baseball team Oakland Athletics, adopted this decision-making structure for picking his players.[46] Baseball

team managers traditionally rely on personal experience, instinct, and the knowledge of professional scouts and agents when choosing players. Billy Beane took a data-driven approach and applied the predictive power of algorithms to assist his decision making by first selecting a small set of potentially suitable players and subsequently verifying these candidates using massive quantities of granular performance data and algorithmic prediction. This approach became so successful that it was soon adopted in other teams and sports, growing into a field now known as sports analytics.

In health care, this structure finds application in AI-based monitoring of bodily functions (e.g., heart rate, temperature, blood pressure) in groups of high-risk patients so as to predict and detect risks of patients developing acute disorders. Because monitoring bodily functions requires the dedicated attention of medical professionals over long periods of time, it is an attractive setting for AI. Deep learning models process anonymized electronic health records and decide which potential emergencies clinicians should attend.[47] In a recent study, researchers used such an approach using AI-based computer vision to monitor patients in an intensive care unit. The system would automatically notify care providers when a patient was experiencing discomfort or had fallen out of bed.[48]

Despite its many potential applications and benefits, this decision-making structure is vulnerable to most of the limitations discussed previously in the full delegation structure. Moreover, the lack of interpretability of the AI-based decision in the second phase can potentially deprive human decision makers from the opportunity to learn from past cases and events.

Aggregated Human–AI Decision-Making Structures

In this structure, decisions—or aspects thereof—are first allocated to human and AI decision makers based on their respective strengths. Human and AI-based decisions are then aggregated into a collective decision using an aggregation rule such as majority voting or (weighted) averaging. In this structure, the AI-based decision maker can be seen as a "member" of the decision-making group, whose decision counts toward the decision outcome. Aggregated decision-making structures can be designed such that human decision makers and AI-based decision makers focus on different or overlapping elements of the decision according to their strengths and weaknesses. In our hiring example, human decision makers may focus on more difficult-to-define factors, such as social fit, and leave it to algorithms to evaluate and predict more objective factors such as productivity—which requires querying specific questions over large amounts of data.

One scenario in which aggregation can be useful concerns decisions taken by investment committees. Consider, for example, Deep Knowledge Ventures (DKV), a Hong Kong based Venture Capital firm focusing on age-related disease drugs and regenerative medicine ventures. DKV formally appointed an algorithm named VITAL (Validating Investment Tool for Advancing Life Sciences) to its board. As the sixth board member, VITAL was given the right to vote on investment decisions. Unlike human board members, VITAL bases its decisions on a

computational analysis of vast amounts of data covering prospective investment companies' financing, clinical trials, intellectual property, and previous funding. Such an analysis involves observing and identifying the role of hundreds of variables and their interactions on investment outcomes and can capture elements of the decision space that are likely to be overlooked by humans. As a case in point, Dmitry Kaminsky, DKV's managing partner, suggests that VITAL has played an important role in helping DKV's board avoid irrational investment decisions in "overhyped projects."[49]

In contrast to hybrid decision-making structures—where there is high interdependence between the human and the AI-based decision maker—this structure allows AI-based and human decisions to be combined independently. In this way, the risk that human decision-making errors and biases are amplified by AI-based decision makers (or vice versa) may be minimized. Moreover, algorithms can find new applications alongside human decision makers to expose biases and errors incorporated in past decisions.[50] Such applications have the potential to turn algorithms into a powerful counterweight to human decision-making errors. Nevertheless, aggregating AI-based decisions with human decisions still exposes organizations to problems of transparency and reliability. For example, in the investment board example above, algorithms can be tweaked so as to output decisions in accordance with the preferences of those with the power to influence the algorithms functioning.

Conclusion

Designing organizational decision-making structures has long been a major concern for managers and organization scholars. The rapid advancement in AI is gradually establishing algorithmic decision makers as key organizational actors. The framework developed here provides a basis for understanding in what ways human and algorithmic decision making can be effectively combined to exploit the advantages of each approach and enable better decisions. This may have the potential to improve organizations if approached with prudence and diligence.

The framework emphasizes that in designing hybrid human–AI decision-making structures, managers should consider the specificity of the decision search space, the interpretability of the decision-making process and outcomes, the size of the alternative set, decision-making speed, and the replicability of decisions. In designing the most appropriate decision-making structure, managers are advised to map these five dimensions to the unique strengths and weaknesses of human and AI-based algorithmic decision making in terms of human's judgment and interpretability and AI's capability of alternative filtering and predicting with high accuracy.

Adding to the more familiar limitations of human decision makers, practitioners and scholars need to advance understanding of the implications of AI's

limitations for organizational decision making. First, there is a risk that AI is "fooled" into altering decision outcomes—either through the manipulation of the data it uses as input or through its design (e.g., by changing weights of predictors). These issues can be difficult to discover due to algorithms' inherent opacity. Thus, inviting algorithmic decision making into organizations will require new regulation and procedures for auditing AI algorithms.[51] Encouraging developments in the AI community will conceivably deliver new techniques for enhancing the robustness and defenses of neural networks against biases and adversarial attacks.[52]

Second, there is by now a vast body of evidence that AI-based decisions amplify human biases in available data. Bias and unfairness embedded in AI decisions are particularly detrimental to vulnerable groups in our society. Countering these grave concerns requires a stronger emphasis on the development of algorithms that can expose biases in data and human decision making, as well as collaboration between the AI community, legal practitioners, policy makers, corporates, and scientists to develop new measures for fair, accountable, and transparent applications of AI in organizations.[53]

Third, introducing AI-based decisions into organizations becomes relatively effective when some level of transparency or interpretability of decisions can be achieved. Managers need to keep abreast of the developments in interpretable and explainable AI.[54] Finally, algorithmic decision-making skills remain highly specialized such that decision outcomes are often difficult to interpret. In introducing AI to organizational decision making, managers must build internal capabilities to decide on the inputs to the algorithm, the algorithms themselves, and the interpretation of predictions. Because AI technologies advance rapidly, organizations must remain vigilant to the strengths and limitations of AI in fully delegated and hybrid human–AI decision-making structures.

Our paper opens up a host of questions for further research. For example, how should performance be evaluated when decisions are partly taken by AI? What are the implications of algorithmic decisions for managerial responsibility? What are the implications of the different decision-making structures presented in our framework for organizational performance? How does the nature of the decision-making context influence the appropriateness of the various approaches? How can concerns regarding trust and accountability be alleviated in a world where AI becomes increasingly important in decision making? and How does the loss of decision-making authority to AI influence the motivation and performance of human decision makers? Addressing these and other questions will make managers and organizational scholars alike, better prepared for an unpredictable future.

Author's Note

The authors contributed equally to this article. The authors thank Michael Haenlein and the three anonymous reviewers for their helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Author Biographies

Yash Raj Shrestha is a senior lecturer and researcher in the Department of Management, Technology, and Economics at ETH Zurich (email: yshrestha@ethz.ch).

Shiko M. Ben-Menahem is a senior lecturer and researcher in the Department of Management, Technology, and Economics at ETH Zurich (email: benmenahem@ethz.ch).

Georg von Krogh is a professor and Chair of Strategic Management and Innovation in the Department of Management, Technology, and Economics at ETH Zurich (email: gvkrogh@ethz.ch).

Notes

1. R. Scott Tindale, Tatsuya Kameda, and Verlin B. Hinsz, "Group Decision Making," in *Sage Handbook of Social Psychology*, ed. Michael A. Hogg and Joel Cooper (Thousand Oaks, CA: Sage, 2003), pp. 381-403.
2. Felipe A. Csaszar and J. P. Eggers, "Organizational Decision-Making: An Information Aggregation View," *Management Science*, 59/10 (October 2013): 2257-2277; Richard M. Cyert and James G. March, *A Behavioral Theory of the Firm* (Englewood Cliffs, NJ: Prentice-Hall, 1963), pp. 169-187; and Herbert A. Simon, *Administrative Behavior* (London, UK: Macmillan 1951).
3. Herbert A. Simon, "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 69/1 (February 1955): 99-118.
4. Ann Langley, Henry Mintzberg, Patricia Pitcher, Elizabeth Posada, and Jan Saint-Macary, "Opening up Decision Making: The View from the Black Stool," *Organization Science*, 6/3 (May/June 1995): 260-279.
5. Georg von Krogh, "Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing," *Academy of Management Discoveries*, 4/4 (2018): 404-409; Georg von Krogh and Johan Roos, *Organizational Epistemology* (New York, NY: St. Martin's Press, 1995); and Michael Christensen and Thorbjørn Knudsen, "Design of Decision-Making Organizations," *Management Science*, 56/1 (January 2010): 71-89.
6. AI has even surpassed human performance in some decision-making contexts, such as game playing (e.g., Chess by DeepBlue and Go by AlphaGo and Deepmind) and visual recognition. See also Ajay K. Agrawal, Joshua S. Gans, and Avi Goldfarb, "Exploring the Impact of Artificial Intelligence: Prediction versus Judgment" (National Bureau of Economic Research Working Paper Series No. 24626, 2018, http://www.nber.org/papers/w24626).
7. For surgery allocation, see Javier Briceño, Manuel Cruz-Ramírez, Martín Prieto, Miguel Navasa, Jorge Ortiz de Urbina, Rafael Orti, Miguel-Ángel Gómez-Bravo, Alejandra Otero, Evaristo Varo, Santiago Tomé, Gerardo Clemente, Rafael Bañares, Rafael Bárcena, Valentín Cuervas-Mons, Guillermo Solórzano, Carmen Vinaixa, Ángel Rubín, Jordi Colmenero, Andrés Valdivieso, Rubén Ciria, César Hervás-Martínez, and Manuel de la Mata, "Use of Artificial Intelligence as an Innovative Donor-Recipient Matching Model for Liver

Transplantation: Results from a Multicenter Spanish Study," *Journal of Hepatology*, 61/5 (November 2014): 1020-1028. For human resource management, see Claire Cain Miller, "Can an Algorithm Hire Better than a Human," *The New York Times*, June 25, 2015, https://www.nytimes.com/2015/06/26/upshot/can-an-algorithm-hire-better-than-a-human.html. For psychological counseling, see Marita Skjuve and Petter Bae Brandtzæg, "Chatbots as a New User Interface for Providing Health Information to Young People," 2018, https://www.nordicom.gu.se/sv/system/tdf/kapitel-pdf/06_bjaalandskjuve_brandtzaeg.pdf?file=1&type=node&id=39926&force. For credit risk prediction, see Vincenzo Pacelli and Michele Azzollini, "An Artificial Neural Network Approach for Credit Risk Management," *Journal of Intelligent Learning Systems and Applications*, 3 (2011): 103-112. For astronomy, see Nicholas M. Ball and Robert J. Brunner, "Data Mining and Machine Learning in Astronomy," *International Journal of Modern Physics D*, 19/7 (2010): 1049-1106. For self-driving cars, see Chenyi Chen Ari Seff, Alain Kornhauser, and Jianxiong Xiao, "Deepdriving: Learning Affordance for Direct Perception in Autonomous Driving," (IEEE international conference on Computer Vision, IEEE Xplore, New York, NY, December 7-13, 2015, https://ieeexplore.ieee.org/document/7410669). For immigration decisions, see Andy Hon Wai Chun, "Using AI for e-Government Automatic Assessment of Immigration Application Forms" (IAAI'07 proceedings of the 19th National Conference on Innovative Applications of Artificial Intelligence, Vancouver, BC, Canada, July 22-26, 2007). For legal counseling, see Cynthia Rudin, "Predictive Policing: Using Machine Learning to Detect Patterns of Crime," *Wired Magazine*, August 2013, https://www.wired.com/insights/2013/08/predictive-policing-using-machine-learning-to-detect-patterns-of-crime/. For hiring, see Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer, "Prediction Policy Problems," *American Economic Review*, 105/5 (May 2015): 491-495.

8. Andreas Kaplan and Michael Haenlein, "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence," *Business Horizons*, 62/1 (January/February 2019): 15-25, See also Erik Brynjolfsson and Tom Mitchell, "What Can Machine Learning Do? Workforce Implications," *Science*, 358/6370 (2017): 1530-1534. Nathan R. Kuncel, David M. Klieger, and Deniz S. Ones, "In Hiring, Algorithms Beat Instinct," *Harvard Business Review*, 92/5 (May 2014): 32.
9. Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli, "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda" (Proceedings of the 2018 CHI conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, April 21-26, 2018).
10. Michael Jordan, "Artificial Intelligence—The Revolution Hasn't Happened Yet," *Medium*, April 19, 2018, https://medium.com/@mijordan3/artificial-intelligence-the-revolution-hasnt-happened-yet-5e1d5812e1e7.
11. Ben Goertzel and Cassio Pennachin, *Artificial General Intelligence* (Berlin, Germany: Springer, 2007).
12. Agrawal et al., op. cit.
13. C. Chet Miller, and R. Duane Ireland, "Intuition in Strategic Decision Making: Friend or Foe in the Fast-Paced 21st Century?" *Academy of Management Perspectives*, 19/1 (February 2005): 19-30.
14. For more details see Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *The Elements of Statistical Learning* (New York, NY: Springer, 2001).
15. Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society*, 3/1 (June 2016): 1-12.
16. Nicolas Papernot, Patrick McDaniel, Somesh Jha, Matt Fredrikson, Z. Berkay Celik, and Ananthram Swami, "The Limitations of Deep Learning in Adversarial Settings" (IEEE European Symposium on Security and Privacy, IEEE Xplore, New York, March 21-24, 2016).
17. Nils Brunsson, *The Organization of Hypocrisy: Talk, Decisions and Actions in Organizations* (New York, NY: John Wiley, 1989).
18. Wilhelm Hofmann, Bertram Gawronski, Tobias Gschwendner, Huy Le, and Manfred Schmitt, "A Meta-Analysis on the Correlation between the Implicit Association Test and Explicit Self-Report Measures," *Personality and Social Psychology Bulletin*, 31/10 (October 2005): 1369-1385. See also Martha S. Feldman and James G. March, "Information in Organizations as Signal and Symbol," *Administrative Science Quarterly*, 26/2 (June 1981): 171-186; John Seely Brown, Allan Collins, and Paul Duguid, "Situated Cognition and the Culture of Learning," *Educational Researcher*, 18/1 (January 1989): 32-42.

19. Sheena S. Iyengar and Mark R. Lepper, "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79/6 (2000): 995-1006.
20. Yoel Inbar, Simona Botti, and Karlene Hanko, "Decision Speed and Choice Regret: When Haste Feels like Waste," *Journal of Experimental Social Psychology*, 47/3 (May 2011): 533-540.
21. Ann Langley, "Between 'Paralysis by Analysis' and 'Extinction by Instinct'," *MIT Sloan Management Review*, 36/3 (Spring 1995): 63.
22. See Jeff Dean, David Patterson, and Cliff Young, "A New Golden Age in Computer Architecture: Empowering the Machine-Learning Revolution," *IEEE Micro*, 38/2 (March/April 2018): 21-29. The time it takes algorithms to learn from past data can be considerable (up to weeks or months), yet the prediction time—the time it takes the algorithm to make a decision—can be near instantaneous.
23. See Martin G. Kocher and Matthias Sutter, "Time Is Money—Time Pressure, Incentives, and the Quality of Decision-Making," *Journal of Economic Behavior & Organization*, 61/3 (November 2006): 375-392; Niv Ahituv, Magid Igbaria, and A. Viem Sella, "The Effects of Time Pressure and Completeness of Information on Decision Making," *Journal of Management Information Systems*, 15/2 (Fall 1998): 153-172.
24. Irvin L. Janis, "Groupthink," *IEEE Engineering Management Review*, 36/1 (2008): 36.
25. See Daniel Kahneman, *Thinking, Fast and Slow* (New York, NY: Farrar, Straus and Giroux, 2011); Andrei Shleifer, "Psychologists at the Gate: A Review of Daniel Kahneman's Thinking, Fast and Slow," *Journal of Economic Literature*, 50/4 (2012): 1080-1091.
26. James F. Smith, Terence R. Mitchell, and Lee Roy Beach, "A Cost-Benefit Mechanism for Selecting Problem-Solving Strategies: Some Extensions and Empirical Tests," *Organizational Behavior and Human Performance*, 29/3 (June 1982): 370-396.
27. For algorithms that involve random components, reproducibility of results can be guaranteed when one keeps track of underlying the "random seeds" used to start off a random number generator.
28. Carsten K. W. De Dreu, Bernard A. Nijstad, and Daan van Knippenberg, "Motivated Information Processing in Group Judgment and Decision Making," *Personality and Social Psychology Review*, 12/1 (February 2008): 22-49. For cognitive biases, see Inbar et al., op. cit.
29. Roy Baumeister, "The Psychology of Irrationality: Why People Make Foolish, Self-Defeating Choices," in *The Psychology of Economic Decisions*, ed. Isabelle Brocas and Juan D. Carrillo, vol. 1 (Oxford, UK: Oxford University Press, 2003), pp. 3-16.
30. Jennifer S. Lerner, Ye Li, Piercarlo Valdesolo, and Karim S. Kassam, "Emotion and Decision Making," *Annual Review of Psychology*, 66 (2015): 799-823.
31. Bart P. Knijnenburg and Martijn C. Willemsen, "Evaluating Recommender Systems with User Experiments," in *Recommender Systems Handbook*, ed. Francesco Ricci (Boston, MA: Springer, 2015), pp. 309-352.
32. Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, "Semantics Derived Automatically from Language Corpora Contain Human-like Biases," *Science*, 356/6334 (April 2017): 183-186.
33. Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, NY: New York University Press, 2018); Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: St. Martin's Press, 2018).
34. Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz, "AI Now Report 2018," AI Now Institute at New York University, 2018, https://ainowinstitute.org/AI_Now_2018_Report.pdf. See also Nick Hopkins, "Britons Make 170,000 Antisemitic Google Searches a Year, Study Finds," *The Guardian*, January 11, 2019, https://www.theguardian.com/news/2019/jan/11/uk-thousands-antisemitic-google-searches-per-year-research.
35. Jack Nicas, "How YouTube Drives People to the Internet's Darkest Corners," *The Wall Street Journal*, February 7, 2018, https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478.
36. See Dimitar Nikolov, Diego F. M. Oliveira, Alessandro Flammini, and Filippo Menczer, "Measuring Online Social Bubbles," *PeerJ Computer Science*, 1 (2015): e38.
37. See Cass R. Sunstein, # *Republic: Divided Democracy in the Age of Social Media* (Princeton, NJ: Princeton University Press, 2018); David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain, "The Science

of Fake News," *Science*, 359/6380 (March 2018): 1094-1096; Efrat Nechushtai and Seth C. Lewis, "What Kind of News Gatekeepers Do We Want Machines to Be? Filter Bubbles, Fragmentation, and the Normative Dimensions of Algorithmic Recommendations," *Computers in Human Behavior*, 90 (January 2019): 298-307.

38. See Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort, "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms" (paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry," a preconference at the 64th annual meeting of the International Communication Association, Seattle, WA, May 22, 2014): Joshua A. Kroll, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu, "Accountable Algorithms," *University of Pennsylvania Law Review*, 165/3 (2016): 633. Also see Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*), https://fatconference.org/;Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai, "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings," *arXiv:1607.06520 [cs.CL]* (2016): 4349-4357.
39. Langley et al., op. cit.
40. Janet A. Sniezek and Timothy Buckley, "Cueing and Cognitive Conflict in Judge-Advisor Decision-Making," *Organizational Behavior and Human Decision Processes*, 62/2 (May 1995): 159-174; Nigel Harvey, Clare Harries, and Ilan Fischer, "Using Advice and Assessing Its Quality," *Organizational Behavior and Human Decision Processes*, 81/2 (May 2000): 252-273.
41. Lars Bo Jeppesen and Karim R. Lakhani, "Marginality and Problem-Solving Effectiveness in Broadcast Search," *Organization Science*, 21/5 (September/October 2010): 1016-1033.
42. Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," *Reuters*, October 9, 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G. For a treatment of unfairness in data-driven decision-making, see Moritz Hardt, "How Big Data Is Unfair," *Medium*, September 26, 2014, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de.
43. Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks," *ProPublica*, May 23, 2016.
44. Evanthia Faliagka, Athanasios Tsakalidis, and Giannis Tzimas, "An Integrated e-Recruitment System for Automated Personality Mining and Applicant Ranking," *Internet Research*, 22/5 (2012): 551-568; Anja Lambrecht and Catherine Tucker, "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," *Management Science*, Published electronically April 10, 2019, doi:10.1287/mnsc.2018.3093.
45. See, for example, David Madras, Toni Pitassi, and Richard Zemel, "Predict Responsibly: Improving Fairness and Accuracy by Learning to Defer," in *Advances in Neural Information Processing Systems, arXiv:1711.06664 [stat.ML]* (2018), pp. 6147-6157; Corinna Cortes, Giulia DeSalvo, and Mehryar Mohri, "Learning with Rejection," in *Algorithmic Learning Theory*, ed. R. Ortner, H. Simon, and S. Zilles, Lecture Notes in Computer Science, vol. 9925 (Cham, Switzerland: Springer, 2016), pp. 67-82.
46. Brad Millington and Rob Millington, "'The Datafication of Everything': Toward a Sociology of Sport and Big Data," *Sociology of Sport Journal*, 32/2 (June 2015): 140-160.
47. Alvin Rajkomar and Eyal Oren, "Deep Learning for Electronic Health Records," *Google AI Blog*, May 8, 2018, https://ai.googleblog.com/2018/05/deep-learning-for-electronic-health.html.
48. S. Yeung, F. Rinaldo, J. Jopling, B. Liu, R. Mehra, N.L. Downing, M. Guo, G.M. Bianconi, A. Alahi, J. Lee, B. Campbell, K. Deru, W. Beninati, L. Fei-Fei, and A. Milstein, "A Computer Vision System for Deep Learning-Based Detection of Patient Mobilization Activities in the ICU," *npj Digital Medicine*, 2 (2019): 11, https://www.nature.com/articles/s41746-019-0087-z.
49. Nicky Burridge, "Artificial Intelligence Gets a Seat in the Boardroom: Hong Kong Venture Capitalist Sees AI Running Asian Companies within 5 Years," *Nikkei Asian Review*, May 10, 2017, https://asia.nikkei.com/Business/Artificial-intelligence-gets-a-seat-in-the-boardroom.
50. Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein, "Discrimination in the Age of Algorithms," *Journal of Legal Analysis*, 10 (2018): laz001, doi:10.1093/jla/laz001.
51. See for example Shaokai Ye, Siyue Wang, Xiao Wang, Bo Yuan, Wujie Wen, and Xue Lin, "Defending DNN Adversarial Attacks with Pruning and Logits Augmentation," (2018 IEEE Global Conference on Signal and Information Processing [GlobalSIP], IEEE Xplore, New

York, NY, February 21, 2019); Sibo Song, Yueru Chen, Ngai-Man Cheung, and C.-C. Jay Kuo, "Defense against Adversarial Attacks with Saak Transform," *arXiv preprint, arXiv:1808.01785 [cs.CV]*, 2018.

52. For recent advancements in reliable AI, see Gagandeep Singh, Timon Gehr, Markus Püschel, and Martin Vechev, "Boosting Robustness Certification of Neural Networks" (ICLR 2019 conference submission, 2018); Gagandeep Singh, Timon Gehr, Matthew Mirman, Markus Püschel, and Martin Vechev, "Fast and Effective Robustness Certification" (NIPS'18 proceedings of the 32nd international conference on Neural Information Processing Systems, Montréal, QC, Canada, December 3-8, 2018).
53. Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review*, 104 (2016): 671; Angwin et al., op. cit.; Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (Proceedings of the 1st conference on Fairness, Accountability and Transparency, *PMLR 81*, 2018, http://proceedings.mlr.press/v81/buolamwini18a.html). Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *Big Data*, 5/2 (2017): 153-163; Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances*, 4/1 (2018): eaao5580.
54. For a recent survey on explainable AI, see Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi, "A Survey of Methods for Explaining Black Box Models," *ACM computing Surveys*, 51/5 (January 2018): Article 93.