TUM School of Management

Essays on Reliability of Intelligent Systems, Cognition in Organization Theory and Digitalization in Financial Accounting

Three Perspectives to Promote Interdisciplinarity in Managerial and Organizational Studies

Dominik G. Fischer

Vollständiger Abdruck der von der promotionsführenden Einrichtung TUM School of Management der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Philipp Maume
1. Prüfender der Dissertation	Prof. Dr. Jürgen Ernstberger
2. Prüfende der Dissertation	Prof. Dr.-Ing. Sanaz Mostaghim

Die Dissertation wurde am 02.07.2020 bei der Technischen Universität München eingereicht und durch die promotionsführende Einrichtung TUM School of Management am 15.11.2020 angenommen.

”How do I know what I think until I see what I say?” (E.M. Forster)

Contents

List of Figures

List of Tables

List of Abbreviations

ACO	Ant Colony Optimization
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area Under the Curve
AS	Ant System
BDSG	Bundesdatenschutzgesetz
cog.	cognitive
env.	environment(al)
EA	Evolutionary Algorithm
EF	Evolved Fitness
F	Fitness
FAIT	Fachausschuss für Informationstechnologie
GA	Genetic Algorithm
GoB	Grundsätze ordnungsgemäßer Buchführung
HGB	Handelsgesetzbuch
HMG	Hidden Markov Gate
HRO	High-Reliability Organization
KI	Künstliche Intelligenz
IDA	Intelligent Distribution Agents
IDW	Institut der Wirtschaftsprüfer
IEC	International Electrotechnical Commission
ISO	Internationale Organisation für Normung
IIT	Integrated Information Theory
IT	Information Technology
LIDA	Learning IDA

LSCC	Largest Strongly Connected Component
MABE	Modular Agent Based Evolver
MB	Markov Brain
NA	not available
NERO	Neuroevolving Robotic Operatives
PublG	Publizitätsgesetz
PH	Prüfungshinweis
PSO	Particle Swarm Optimization
PS	Prüfungsstandard
RFID	Radio-Frequency Identification
rtNEAT	real-time Neuroevolution of Augmenting Topologies
RS	Stellungnahme zur Rechnungslegung
SEM	Standard Error of the Mean
TPM	Transition Probability Matrix
TF	Task Fitness
WPK	Wirtschaftsprüferkammer
WPO	Wirtschaftsprüferordnung
ZUGFeRD	Zentraler User Guide des Forums elektronische Rechnung Deutschland
d.h.	das heißt
et al.	et alii, et aliae
e.g.	exempli gratia
i.e.	id est
n.F.	neue Fassung
ref.	reference
sog.	sogenannt
Tz.	Teilziffer/Textziffer
vgl.	vergleiche
z.B.	zum Beispiel

Chapter 1

Preface

1.1 Motivation and scope

1.1.1 Interdisciplinarity in organization studies

This dissertation presents four essays to push frontiers in organization studies and to explain the challenges of organizational change during the digitalization. The four articles draw on methods and theory from the fields of cognition and swarm intelligence to investigate the relationship between individual behavior and organizational performance. For instance, a particular challenge in organization studies is that little is known about the connection between the organization and its members[1] (Barney & Felin 2013). I use an interdisciplinary[2] approach to investigate the dynamics between individual level and organizational level observations, since advancing knowledge about this gap is hardly feasible with traditional methods (Salvato & Rerup 2011).

An interdisciplinary approach to advance organizational theory might be *"radical"* for some readers (Nadkarni et al. 2018), especially when referring to cognitive science for finding solutions to bridge observations on the individual and the organizational levels. However, cognition, a discipline that tries to understand the computational processes of the mind, faces similar macro-micro challenges. For instance, when considering the brain's architecture, billions of neurons interact simultaneously and influence the human's mental states. Neuroscience tries to advance knowledge on how the neuron's micro-level activity correlates with the mental state on the macro-level (Thagard 2017). Cognitive science is an interdisciplinary discipline involving and influencing philosophy, linguistics, anthropology, neuroscience, computer science, and psychology (Miller 2003). In general, *"interdisciplinary research ... can be one of the most productive and inspiring of human pursuits – one that provides a format for conversations and connections that lead to new knowledge"* (National Academy of Sciences 2004). However, *borrowing* theory from diverse disciplines has boundaries and can create misconceptions (Whetten et al. 2009). Therefore, I briefly address the characteristics of interdisciplinary research to prepare the readers for a broader understanding of this topic and the essay itself.

Interdisciplinarity is defined as *"the quality or fact of involving different areas of knowledge or study"* (Oxford University Press 2020) *"to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice"* (National Academy of Sciences 2004). Elaborating on knowledge and using methods from different disciplines is a matter in each of the dissertation's Chapters: Wether using experimental computer simulations to investigate reliability (see Chapter 2 and Chapter 3), neuroscientific theory to explain the collective mind in organizations (see Chapter 4), or connecting financial accounting and major trends in digitalization (see Chapter 5).

Computational simulation offers a promising way to introduce methods in organization studies, which can support novel approaches in macro-micro-level research (Thau et al. 2014, Harrison et al. 2007). Specifically, agent-based modeling and simulation help to better understand the behavior of organizations and their members

[1] Also known as *micro-macro-links*.

[2] In particular, I build on literature in organizational studies, computational biology, and cognition.

(Fioretti 2013) and allow for manipulating variables, which might be difficult in field or laboratory experiments, e.g., the organization's size or the individual intelligence and skill set. Further, using computer simulations, it is convenient to collect and analyze data on multiple levels of analysis simultaneously. Doing so offers methods to advance knowledge in organization studies and provide new ways of theorizing.

I argue that experimental simulation studies can be a proper method to explain how interrelating individual behavior can emerge mental states on a collective level. In Chapter 4, I use a neuroscientific theory (Oizumi et al. 2014), that provides an algorithmic framework (Albantakis et al. 2014, Mayner et al. 2018), to explain the physical structure of mental states in an organization. The last Chapter about digitalization in auditing shows an alternative perspective on interdisciplinarity. It serves as a useful narrative to explain current issues in traditional organizations during the digital transformation. That practice-related discussion shall be a matter of the reflection in Chapter 6. Doing so also shows how it is possible to adapt the abstract and theoretical work on cognition in organizational studies to support the practice, e.g., during strategical decision-making.

1.1.2 Uncertainty in the age of digitalizations

From a much broader perspective, in the age of digitalization, we need to understand the connections between micro- and macro-levels: Machines and employees in contrast to the organizations. A lack of knowledge of the interacting processes might decrease organizations' robustness during uncertain change, which might be due to the following reasons: First, the pace of technological innovations and new business models is increasing (Salovaara et al. 2019). Second, an illustration of a digitalized industry is that every human and every machine is interconnected, and everything can communicate with each other. Therefore, small incidents can cause a chain reaction that leads to significant system-wide crises.[3] Third, it is apparent that the focus is currently on the implementations of technological innovation, while there is a lack of supporting employees and acquiring talents or even a clear digitalization strategy (Salovaara et al. 2019, Tabrizi et al. 2019).

High-Reliability Organizations (HROs) can be models to find practices supporting reliability in organizations. HROs are organizations, which perform nearly error-free in highly uncertain environments. Traditional examples are aircraft carriers, medical intensive care units, or wildfire fighters (Weick & Roberts 1993). There is already a growing interest in applying HRO practices in information technology systems (Dernbecher et al. 2014, Carlo et al. 2012, Aversa et al. 2018). In general, they suggest, that there can be overconfidence in decision-support systems with Artificial Intelligence (AI), that organizations need to be more reliable even if algorithms can be hardly adapted to environmental change and that the careless implementation of Information Technology (IT) can even harm the reliability of the organization. These findings show the relevance of HRO research for organizations in the age of digitalization. The other way around, there is a potential for the literature of organization studies when investigating the interaction between individuals and AI systems, too (von Krogh 2018).

For an organization to stay reliable, it depends on its members' mindful and smart behavior (Weick & Roberts 1993). In that scope, to be reliable means that the organization remains healthy and robust even though unlikely or unpredictable situations appear. Due to increasing economic and ecological disruptions, it becomes more important to develop strategies to ensure the organization's reliability. Even if there is a high potential for investing in digital innovations, there is no way around the training and recruitment of highly skilled employees and a dedicated digitalization strategy (Carpi et al. 2019, Valorinta 2009, Muhren et al. 2007). This fact has an essential relation to HROs: Without technology, HROs could not be successful or reliable.[4] Still, they need people with high situational awareness, professional experience, and problem-solving skills (Weick & Roberts 1993). Hence, through investigating HROs, it is possible to gain knowledge about such essential skills and how it would be possible to implement them into practice (Gebauer 2017), e.g., for the digitalization in organizations.

Nevertheless, the number of automated processes will increase rapidly. Simultaneously, highly-automated organizations require employees, to observe the complex automated systems since algorithms lack in detecting

[3] E.g., in the 2010 Flash Crash, trading algorithms were strongly involved in a temporary trillion dollar stock market crash (Kirilengo et al. 2017).

[4] E.g., a nuclear power plant would be unimaginable without automated monitoring systems.

unexpected change (Carpi et al. 2019). Those positions would require skills, which are difficult, if not impossible, to automate, e.g., the attention for a complex environment, interacting with people from diverse fields, creativity for problem-solving, or sharing knowledge and inspirations to have imagination or empathy (Maslach et al. 2017, Carpi et al. 2019, Precht 2018). HROs can teach us to understand how to prepare organizations for these challenges.

1.1.3 Cognition in organizational studies

In the first three essays, I apply cognitive science to explain the processes and behavior of organizations and their members. This research contributes to literature in organizational cognition, a subsection of organization studies (e.g., Walsh 1995, Sutcliffe et al. 2016). I briefly introduce this scope and their key areas of interest to underline the motivation for my contribution.

Organizational cognition is not a traditional topic, yet already well established since the 1990s (Walsh 1995). The Academy of Management, for example, provides special interest groups for *"Managerial and Organizational Cognition" (MOC)* and *"Organizational Neuroscience"* to support the researchers. The literature in this scopes is diverse and reaches from performing experiments with brain activity measures during financial negotiations to applying cognitive science to theorize about organizational behavior. Generally, using neuroscientific and psychological knowledge in management helps to understand essential phenomena (Hodgkinson & Healey 2008). For instance, scholarly concepts from that literature are bounded rationality (Simon 1947, March & Simon 1958), tacit knowledge (Polanyi 1962, Nonaka & Nonaka 1994), or sense-making (Weick et al. 2005, Kudesia 2017).

Nowadays, cognition is an omnipresent term in research. AI is called to be cognitive (Gamez 2008), cognitive economics is *"the economics of what is in people's minds"* (Kimball 2015, p. 2), and behavioral finance applies cognitive psychology (Ritter 2003). The scattering of the term *cognition* across the literature makes it hard to understand its nature, and often it is reduced to neuroscientific methods only. That is why I provide a clear definition of cognition to understand its purpose better:

The fundamental interest of cognition is to find a *computational-representational understanding of the mind (CRUM)* (Long 2005). In cognition, the most basic assumption is that the mind is thinking and that thoughts matter for our behavior. Cognition is mostly related to the methods of neuroscience, which try to explain our mind's processes by investigating the activity of the neurons in the brain (*connectionism*). But there are further approaches to model the computational processes and representations, including *logic, rules, concepts, analogies*, or *images* (Long 2005).

Cognition is a rather new field of research and was trending in the mid-twentieth century (Miller 2003). When cognition was established as a scientific discipline, there was a dominance of behaviorism. In behaviorism, the researcher tries to understand the subject's behavior, which results from a specific reflex. An example of behavioral analysis would be to condition employees to receive a bonus when their performance is outstanding. The assumption would be that this positive incentive increases work motivation. By definition, behaviorists would not be interested in the processes, which lead from a specific stimulus to an appropriate reflex (Graham 2019). Following behaviorism, it is possible to study the behavior of organizations. However, it becomes challenging to simultaneously identify the internal processes of organizations that execute the behavior on a higher level, e.g., the investigator risks biases and aggregation errors (Rousseau 1985). By introducing cognition, organizational scholars can refer to theories and methods to study the organization's internal processes and refer them to the organizational level. Example topics for such intentions are the concepts of distributed knowledge structures (Hutchins 1995), organizations evolving (Strang & Aldrich 2002), or the process of collective minding (Weick et al. 2008).

A critique on organizational cognition is that the scope suffers from the argument that mental processes on an organizational level are only anthropomorphisms, even though it spawned some important managerial concepts (Walsh & Ungson 1991, Jones 1995). While using narratives can be a powerful tool to build theory in organization studies (Shapira 2011, Weick 1989), some scholars criticize this approach's dominance and call for more evidence-based research. I contribute to this discussion by using concepts of organizational cognition and validating them with contemporary research in cognition. In Chapter 4, I use the concept of organizational mindfulness to visualize a conceptual framework for measuring the organization's collective mind. Organizational mindfulness originates

from research in HROs and explains why such organizations stay reliable in challenging situations. In that context, mindfulness represents a particular mental state with a high level of awareness and attention (Kabat-Zinn 1996).[5] Karl Weick and colleagues assume that, under the right conditions, organizations can be mindful, as well as humans can be, which qualifies them to detect endangering situations and react appropriately to them (Weick et al. 2008, Weick & Roberts 1993).

In brief, organizational mindfulness suggests that reliable organizations have processes to maintain qualities of *anticipation* and *resilience*:

Anticipation describes that the organization is prepared for uncertain events, e.g., events like the COVID-19 outbreak. They are not or not only focused on their success but also on situations, which can lead to failure or crisis, and prepare for them. Mindful organizations tend not to simplify operations. Simplification involves that small anomalies can be overseen, but those anomalies can pile up and cause an incident. Further, mindful organizations are sensitive to their operations and avoid narrow-minded perspectives. That involves that the big picture of the organization and its position in the environment is shared across the organization's members (Weick et al. 2008).

Resilience describes how well and fast an organization recovers from endangering situations (Weick et al. 2008). Here organizations must show commitment to resilience. In critical situations, members of the organization would not stick to their usual work schedule, but *do what is necessary for the organization to recover*. A consequence of this commitment is that the members are willing to temporarily break up hierarchies and get comfortable or even be positive within the unexpected situation, e.g., to be able to *think outside the box* for a problem solution (Fraher et al. 2017).

From a structural perspective, a mindful organization only exists in a dynamic process. Since their physical substrates would be individual behavior, the collective mind in organizations can only be present in action (Sandelands & Stablein 1987, Weick & Roberts 1993). Further, individual behaviors have to be interrelating with other individuals' behavior to create an integrated network.[6] This kind of tight coupling and interrelating creates redundant operational structures but increases the reliability of the organization itself.

1.1.4 The brain, mind and consciousness

After introducing the core topics of this dissertation, it is necessary to distinguish between the brain, mind, and consciousness. Although these concepts are related, it is not easy to differentiate them. Their cultural use is probably a significant reason (Knobe & Prinz 2008, Arico et al. 2011), e.g., by using proverbs in a daily language like: *"Where is my mind?"* or *"Use your brain!"* However, in this dissertation, the differences between the brain, the mind, and consciousness should be more transparent, because they are part of discussions on both the individual and collective level.

A brain is mainly an anatomic object, part of the central nervous system, and located within the skull. The function of the brain is to receive the body's and the environment's stimuli and transform it into an intelligent reaction (Encyclopædia Britannica 2020). Intelligent behavior is mainly derived from past experiences. It is possible to investigate the functions of the brain by neuroscientific methods (Long 2005).[7] Due to such methods, the brain is sometimes confused or equated with the mind, but they must be kept separate: While the brain is an objective phenomenon, the mind is a subjective idea (Block 1995).

Cognitive neuroscience tries to explain how the complex processes in the brain enable the mind. However, there are diverse theories – not a distinct definition – about the mind itself.[8] In very general terms, the mind is *"the complex of faculties involved in perceiving, remembering, considering, evaluating, and deciding."* (Encyclopædia Britannica 2016). The central assumption in this cognitive definition is that the mind is capable of thought: The mind is the entity, which is thinking and establishes a person's identity, a persona, provides the ability for introspection and to control the conscious thoughts. While in the early days of psychology, it was assumed that everything in mind is perceived consciously, nowadays, literature distinguishes between the *unconscious* and *conscious* mind

[5] Chapter 4 offers a detailed delimitation between the terms *organizational mind* and *organizational mindfulness*.
[6] See *loose* and *tight* coupling (Weick 1976).
[7] E.g., *Electroencephalography, Functional magnetic resonance imaging* or *magnetic resonance imaging*.
[8] See the philosophical discussions about the *theory of mind*.

(Baars 1997*b*). A useful analogy to differentiate the concepts is to view the brain as the hardware of the computer, while the mind would be the software that is controlling the computer (Block 1995, Searle 1980).

Consciousness can be defined as the perception of our current experience without any valuation. Analogically, if the stage of a theatre would be the mind, there would be a spotlight representing the attention, and the conscious entity would observe the highlighted scene on the stage (Baars 1997*a*). Even if the stage would be empty, and all spotlights would be off, there is still consciousness experiencing a pitch-black scene (Tononi 2004). Hence, consciousness might be distinct from the analytical thought (Oizumi et al. 2014), and the quality of perception would be sufficient for having conscious experience about the mind. This logic implies that there is a difference between a thinking entity or an entity that is aware of its thoughts.

I use the Integrated Information Theory (IIT) to measure the complexity of artificial brains[9] and the behavioral structure of organizations. IIT is a neuroscientific theoretical construct to explain the quality and quantity of conscious experience within the mind (Oizumi et al. 2014, Tononi 2004, Tononi et al. 2016). IIT is the most comprehensive theory to measure the subjectivity of mental states[10] (Kyumin Moon & Hongju Pae 2019), which might be useful to assess if coma patients are still conscious or brain-dead (Koch et al. 2016). IIT's algorithmic framework can also be used as an information-theoretic complexity measure for dynamic and interrelated state-dependent systems. Since the theory is not isolated to the human mind, it can be modeled on each other physical system. This abstraction enables the usage of IIT to measure the mind of organizations, e.g., Schwitzgebel (2015). Doing so requires to use IIT's axiomatic system (Oizumi et al. 2014) and postulate it on the characteristics of organizations. The abstract orientation of IIT made it possible to measure the complexity of the artificial organisms' brain[11] and the collective mind or organizations[12], too.

1.2 Structure of the dissertation and main findings

1.2.1 Reliability of simulated animats in groups

Chapter 2 and Chapter 3 cover the findings of an extensive simulation study. Both are motivated by the questions about which variables influence the reliability of organized animats. Animats are simple agents with cognitive abilities similar to organisms, e.g., ants or bees (Edlund et al. 2011, Albantakis et al. 2014). The simulation study has an abstract character since little is known in this field. We simulate the evolution of organized animal swarms navigating between two rooms without colliding against each other. These studies help to advance knowledge about the optimal size of organizations and how the organization's performance depends on the abilities of the individual members.

Simulating agents acting in an organized group is already a common practice. Not only to study swarm behavior but also to solve optimization problems (e.g., Asgari et al. 2016, Dorigo et al. 2006, Garnier et al. 2014). Still, little is known about how the size of the group, the cognitive abilities of the agent, or uncertain environmental conditions influence behavior (Pinter-Wollman et al. 2018). Consequently, depending on the internal and external constraints, the behavior of agents and their ability to cooperate varies.

In the simulation experiments, we are wondering about the specific size of the organization in terms of the number of individual members. As a second dimension, we manipulate the individual's potential skills in terms of memory, sensory capacity, or motoric abilities during evolution. After evolving the animats, we expose them with uncertain conditions like changing the group size or changing the static environment (Fischer et al. 2020). Since the environment has constant space, changing the group size varies the environment's density, and the animats are less or more likely to meet group members. Changing the static environment makes it easier or harder to navigate between the rooms, e.g., by adding or removing walls. We argue that those post-evolutionary tests provide data to assess the reliability of animats under uncertain conditions. The abstract nature of our study can be useful to

[9] The artificial brains, here Markov brains, can be evolved using genetic algorithms (Hintze et al. 2017, Clifford Bohm, Nitash C. G. 2017).
[10] In this dissertation I used the software *PyPhi* to measure the complexity of the mental states (Albantakis et al. 2014, Mayner et al. 2018).
[11] See Chapter 2 and Chapter 3.
[12] See Chapter 4.

theorize about the dynamics between the organizational and individual levels. Further, it sheds light on the high complexity of studies investigating multiple levels of analysis.

The findings are divided into the results for the evolution of the animats and the results of the reliability assessment: First, we find that it is dependent on the group size for animats to evolve flexible behavior, reasonable task fitness, and higher brain complexity. For our experiment design, the group size must be neither too small or too large. Second, we find that there can be a relationship between the group size an animat evolves in and the corresponding reliability. In our design, an animat who evolves in a balanced group size shows the best reliability under uncertain conditions. Further, we can show a correlation between high brain complexity, the group size, and optimized sensor capacity.

The evolutionary simulation experiments using a Genetic Algorithm (GA) to generate collective behavior. This method visualizes a way to study reliability in organized groups while being able to control conditions and collect data on multiple levels. In general, the studies show that the group size can have an effect on reliability in uncertain conditions and that rather general cognitive abilities, which relate to heuristics, can avoid over-specialization.

Chapter 2 and Chapter 3 have a focus on computational biology, but offer insights and inspiration for organizational studies, too. The biological nature of the studies is mainly due to the high abstraction of the problem formulation, which is necessary when relating the cognitive processes of individuals to the behavior of the organized group. The results of this study can serve to design more specific experimental studies to investigate the dynamics between the organization and its members and their relationship to the reliability in uncertain environments.

1.2.2 Measuring the collective mind

Chapter 4 offers a conceptual framework about the collective mind in organizations and a theory alignment to measure that concept. Discussions about the idea of a supra-individual mind are not novel (Walsh 1995) and date back to Greek philosophy (Voegelin 2000). Sociological studies have picked up this idea again to describe social behavior and solidity (Durkheim 1895). Such discussions are the foundations for theories in organizational cognition: Assuming a mental state in an organization, organizational mindfulness, implies that there is a mind causing the mindful state, e.g., expressed by the quality of attention and awareness. Hence, we would need to find a clear definition of the collective mind in organizations to understand the organization as a cognitive system fully. Independently from this logical reasoning, the literature is uncertain if the collective mind is only an anthropomorphism or if it can be an observable phenomenon, too (Weick & Roberts 1993, Jones 1995).

I want to offer a view of the collective mind in organizations as an objective phenomenon, while metaphors and anthropomorphisms have a long tradition in organization studies (Cornelissen 2006). The study reveals and explains the gaps in the current conceptualization of a collective mind in organizations. Further, I offer a conceptual framework of the collective mind in organizations, which is based on interrelating behavior as the mind's physical substrates (Durkheim 1895). The physical substrate would be the smallest building block. The central assumption to enable a collective mind is that individual behavior has to be integrated with the behavior of other organizational members in a smart fashion (Sandelands & Stablein 1987, Weick & Roberts 1993). This behavioral network can be used to apply neuroscientific theory to investigate the dynamics on a collective level. Therefore, I blend in IIT, which describes the conscious experience of the mind to explain how this behavior needs to be structured to give rise to mental states.

This theoretical elaboration has, in particular, three contributions to the literature: First, it offers a model to validate the existence of a collective mind in an organization. Even if a mental state in an organization would be very dynamic and volatile, it can be a supra-individual phenomenon, irreducible to its members (Sandelands & Stablein 1987, Weick 1989, March 1962). This idea can be confirmed by recent neuroscientific theory (e.g., Koch et al. 2016). Second, the adaptation of IIT's axioms enables the application of algorithms to measure the complexity of the mind's current mental state, which offers a method to validate if an organization can have mental states or not and what quality it would have. Still, the complexity of an collective mind in organizations is rather trivial to that of a human mind. However, even understanding simple *mind-like processes* can help understanding

organizations better (Sandelands & Stablein 1987). Third, I highlight to distinguish between the collective mind and organizational performance. This elaboration helps to improve the study of their relations and would help to study concepts like organizational mindfulness or organizational memory.

Generally, I find that not every organization would qualify to give rise to a collective mind. Assuming that IIT holds, an organization would need specific characteristics like a complex interdependency of individual behavior. If such a structure were given, the mind would still be a dynamic and volatile phenomenon. The presented study argues that the collective mind is more than a metaphor and allows us to scientifically study its cognitive processes. Future studies can use this approach to quantify knowledge in organizations or to detect very robust behavioral networks.

1.2.3 Digitalization in the audit industry

Chapter 5 contains the last essay of this dissertation. The process of digitalization in the external audit will deal as a narrative to discuss possible pathways to implement the theoretical insights into practice. This case is qualified since it reflects the clash of a conservative industry in light of the uncertainty in digitalization. The actual essay contains an evaluation of possible trends, risks, and consequences of digitalization – as a technological phenomenon – in the auditing. Auditing firms are notably different from modern startups: Auditors' professional obligations[13] make the industry less flexible and dynamic. While startups aim to be disruptive, scalable, and to develop flexible business models, auditors face a relatively static market, high expenses for innovations[14], and at the management level a low willingness to change. Further, the clients of audit firms are forced by law to hire an external auditor[15], and primarily, the audited firms want to receive an audit opinion for their annual report. Auditors are only limited to provide further services because otherwise, they would disqualify themselves since they are not independent of the firm anymore (Richter 1977) since this independence is part of the professional obligations[16].

The goal of this essay is to evaluate how the age of digitalization influences this duty. We identified two significant problems concerning this specific industry: First, the cost for innovation and change is rather high – the causes are the strict laws and professional obligations. Hence it is expensive to develop innovations that fulfill the legal obligations, e.g., such as confidentiality. Second, the profession is rather conservative when it comes to change, and neither their clients nor the law forces them to improve their operating processes. Those are, to some extent, predefined by national institutions of auditors. For instance, in Germany, the Institut der Wirtschaftsprüfer (IDW) publishes regulations, that contain audit standards and references, which comply with the legal guidelines. For a significant change to happen, the regulator needs to update the legal standards that would require the auditors to change their processes.

I suggest that organizational mindfulness offers practices, which could help the industry to prepare themselves in an organic, more natural way. It is not the technology that counts in the digital future, but the people who apply it and, apart from all, the correct digitalization strategy (Salovaara et al. 2019, Tabrizi et al. 2019). Practices of *anticipation* and *resilience* can shape awareness for problems, which can be solved with technology and issues that should first be addressed through human intervention (Gebauer 2017). A more in-depth explanation of that idea will be part of the reflection in the last Chapter.

[13] E.g., see § 43 WPO.
[14] Due to strong regulation (see Section 5.2.3).
[15] E.g., see § 316 HGB.
[16] See § 43(1) WPO.

1.3 References

Albantakis, L., Hintze, A., Koch, C., Adami, C. & Tononi, G. (2014), 'Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity', *PLoS Computational Biology* **10**(12), e1003966.
URL: *https://dx.plos.org/10.1371/journal.pcbi.1003966*

Arico, A., Fiala, B., Goldbert, R. F. & Nichols, S. (2011), 'The Folk Psychology of Consciousness', *Mind & Language* **26**(3), 327–352.

Asgari, A., Hassani, K. & Lee, W.-s. (2016), 'Simulating collective intelligence of bio-inspired competing agents', *Expert Systems With Applications* **56**, 256–267.
URL: *http://dx.doi.org/10.1016/j.eswa.2016.03.016*

Aversa, P., Cabantous, L. & Haefliger, S. (2018), 'When decision support systems fail: Insights for strategic information systems from Formula 1', *The Journal of Strategic Information Systems* **27**(3), 221–236.

Baars, B. J. (1997*a*), *In the Theater of Consciousness*, number 4, Oxford University Press.

Baars, B. J. (1997*b*), 'Some Essential Differences between Consciousness and Attention, Perception, and Working Memory', *Consciousness and Cognition* **6**(2-3), 363–371.

Barney, J. & Felin, T. (2013), 'What are Microfoundations?', *Academy of Management Review* **27**(2), 138–155.

Block, N. (1995), The mind as the software of the brain, *in* S. S. D. N. Osherson, L. Gleitman, S. M. Kosslyn, S. Smith, ed., 'An Invitation to Cognitive Science', MIT Press, pp. 1–16.

Carlo, J. L., Lyytinen, K. & Boland, R. J. (2012), 'Dialectics of Collective Minding: Contradictory Appropriations of Information Technology in a High-Risk Project', *MIS Quarterly* **36**(4), 1081–1108.

Carpi, R., Claus, P., Mattik, I. & Schulze, P. (2019), 'What high-reliability organizations get right'.
URL: *https://www.mckinsey.com/business-functions/operations/our-insights/what-high-reliability-organizations-get-right*

Clifford Bohm, Nitash C. G., A. H. (2017), MABE (Modular Agent Based Evolver): A framework for digital evolution research, *in* 'Proceedings of the European Conference on Artificial Life', MIT Press, pp. 76–83.

Cornelissen, J. P. (2006), 'Making Sense of Theory Construction: Metaphor and Disciplined Imagination', *Organization Studies* **27**(11), 1579–1597.

Dernbecher, S., Risius, M. & Beck, R. (2014), 'Bridging the Gap – Organizational Mindfulness and Mindful Organizing in Mobile Work Environments', *Ecis* pp. 1–16.

Dorigo, M., Birattari, M. & Stutzle, T. (2006), 'Ant colony optimization', *IEEE Computational Intelligence Magazine* **1**(4), 28–39.

Durkheim, E. (1895), *The Rules of Sociological Method*.

Edlund, J. A., Chaumont, N., Hintze, A., Koch, C., Tononi, G. & Adami, C. (2011), 'Integrated Information Increases with Fitness in the Evolution of Animats', *PLoS Computational Biology* **7**(10), e1002236.

Encyclopædia Britannica (2016), Mind, *in* 'Encyclopædia Britannica'.
URL: *https://www.britannica.com/topic/mind*

Encyclopædia Britannica (2020), Brain, *in* 'Encyclopædia Britannica'.
URL: *https://www.britannica.com/science/brain*

Fioretti, G. (2013), 'Agent-Based Simulation Models in Organization Science', *Organizational Research Methods* **16**(2), 227–242.

Fischer, D., Mostaghim, S. & Albantakis, L. (2020), 'How cognitive and environmental constraints influence the reliability of simulated animats in groups', *PLOS ONE* **15**(2), e0228879.

Fraher, A. L., Branicki, L. J. & Grint, K. (2017), 'Discovering How U.S. Navy SEALs Build Capacity for Mindfulness in High-Reliability Organizations (HROs)', *Academy of Management Discoveries* **3**(3), 239–261.

Gamez, D. (2008), 'Progress in machine consciousness', *Consciousness and Cognition* **17**(3), 887–910.

Garnier, S., Hamann, H., Montes, M., Christine, D. O., Eds, T. S. & Hutchison, D. (2014), Swarm Intelligence, *in* Gerhard Goos, J. Hartmanis & J. van Leeuwen, eds, 'LNCS 8667', Brussels.

Gebauer, A. (2017), *Kollektive Achtsamkeit organisieren*, Schäffer Poeschl.

Graham, G. (2019), Behaviorism, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2019 edn, Metaphysics Research Lab, Stanford University.

Harrison, R. J., Zhiang, L., Glenn, C. R. & Carley, K. M. (2007), 'Simulation modeling in organizational and management research', *Academy of Management Review* **32**(4), 1229–1245.

Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., Goldsby, H., Bohm, C. & Adami, C. (2017), 'Markov Brains: A Technical Introduction'.
URL: *http://arxiv.org/abs/1709.05601*

Hodgkinson, G. P. & Healey, M. P. (2008), 'Cognition in Organizations', *Annual Review of Psychology* **59**(1), 387–417.
URL: *http://www.annualreviews.org/doi/10.1146/annurev.psych.59.103006.093612*

Hutchins, E. (1995), 'How a Cockpit Remembers its Speed', *Cognitive Science* **19**, 265–288.

Jones, M. (1995), 'Organisational learning: Collective mind or cognitivist metaphor?', *Accounting, Management and Information Technologies* **5**(1), 61–77.

Kabat-Zinn, J. (1996), *Mindfulness Meditation*.

Kimball, M. (2015), 'Cognitive Economics', *Japanese Economic Review* **66**(2), 167–181.
URL: *http://doi.wiley.com/10.1111/jere.12070*

Kirilengo, A., Kyle, A. S., Samadi, M. & Tuzun, T. (2017), 'The Flash Crash: High-Frequency Trading in an Electronic Market', *The Journal of Finance* **72**(3), 967–998.
URL: *http://doi.wiley.com/10.1111/jofi.12498*

Knobe, J. & Prinz, J. (2008), 'Intuitions about consciousness: Experimental studies', *Phenomenology and the Cognitive Sciences* **7**(1), 67–83.

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016), 'Neural correlates of consciousness: progress and problems', *Nature Reviews Neuroscience* **17**(5), 307–321.

Kudesia, R. S. (2017), Organizational Sensemaking, *in* 'Oxford Research Encyclopedia of Psychology', Vol. 53, Oxford University Press, pp. 286–305.

Kyumin Moon & Hongju Pae (2019), 'Making Sense of Consciousness as Integrated Information: Evolution and Issues of Integrated Information Theory', *Journal of Cognitive Science* **20**(1), 1–52.

Long, D. M. (2005), *Mind – Introduction to Cognitive Science*, Vol. 15.

March, J. G. (1962), 'The Business Firm as a Political Coalition', *The Journal of Politics* **24**(4), 662–678.

March, J. G. & Simon, H. A. (1958), *Organizations*, Wiley, New York, NY.

Maslach, D., Branzei, O., Rerup, C. & Zbaracki, M. J. (2017), 'Noise as Signal in Learning from Rare Events', (September).

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R. & Tononi, G. (2018), 'PyPhi: A toolbox for integrated information theory', *PLOS Computational Biology* **14**(7), e1006343.

Miller, G. A. (2003), 'The cognitive revolution: A historical perspective', *Trends in Cognitive Sciences* **7**(3), 141–144.

Muhren, W. J., Van, G., Eede, D., Van De Walle, B. & Muhren, W. (2007), 'Organizational Learning for the Incident Management Process', *European conference on information systems.* pp. 576–587.

Nadkarni, S., Gruber, M., DeCelles, K., Connelly, B. & Baer, M. (2018), 'New Ways of Seeing: Radical Theorizing', *Academy of Management Journal* **61**(2), 371–377.
URL: *http://journals.aom.org/doi/10.5465/amj.2018.4002*

National Academy of Sciences (2004), Executive Summary, *in* 'Facilitating Interdisciplinary Research', National Academies Press, Washington, D.C., pp. 1–15.

Nonaka, I. & Nonaka, I. (1994), 'A Dynamic Theory of Organizational Knowledge Creation', *Organization Science* **5**(1), 14–37.

Oizumi, M., Albantakis, L. & Tononi, G. (2014), 'From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* **10**(5), 1–25.

Oxford University Press (2020), Interdisciplinarity, *in* 'Oxford Lerner's Dictionary'.
URL: *https://www.oxfordlearnersdictionaries.com/definition/english/interdisciplinarity*

Pinter-Wollman, N., Penn, A., Theraulaz, G. & Fiore, S. M. (2018), 'Interdisciplinary approaches for uncovering the impacts of architecture on collective behaviour', *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1753).

Polanyi, M. (1962), 'Tacit Knowing', *Reviews of Modern Physics* **34**(4), 601–616.

Precht, R. D. (2018), *Jäger, Hirten, Kritiker: Eine Utopie für die digitale Gesellschaft*, Goldmann Verlag.

Richter, M. (1977), 'Die inkompatibilität von jahresabschlußprüfung und unternehmensberatungen durch wirtschaftsprüfer', *Journal für Betriebswirtschaft* **27**(1), 21–42.

Ritter, J. R. (2003), 'Behavioral finance', *Pacific-Basin finance journal* **11**(4), 429–437.

Rousseau, D. M. (1985), 'Issues of level in organizational research: Multi-level and cross-level perspectives', *Research in Organizational Behavior* **7**, 1–37.

Salovaara, A., Lyytinen, K. & Penttinen, E. (2019), 'High reliability in digital organizing: Mindlessness, the frame problem, and digital operations', *MIS Quarterly: Management Information Systems* **43**(2), 555–578.

Salvato, C. & Rerup, C. (2011), 'Beyond collective entities: Multilevel research on organizational routines and capabilities', *Journal of Management* **37**(2), 468–490.

Sandelands, L. E. & Stablein, R. E. (1987), 'The Concept of Organization Mind', *Research in the Sociology of Organizations* **5**, 135–161.

Schwitzgebel, E. (2015), 'If Materialism Is True, the United States Is Probably Conscious', *Philosophical Studies* **172**, 1697–1721.

Searle, J. R. (1980), 'Minds, brains, and programs', *Behavioral and Brain Sciences* **3**, 417–424.

Shapira, Z. (2011), '"I've Got a Theory Paper—Do You?": Conceptual, Empirical, and Theoretical Contributions to Knowledge in the Organizational Sciences', *Organization Science* **22**(5), 1312–1321.

Simon, H. A. (1947), *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, New York: Macmillan.

Strang, D. & Aldrich, H. (2002), 'Organizations Evolving'.

Sutcliffe, K. M., Vogus, T. J. & Dane, E. (2016), 'Mindfulness in Organizations: A Cross-Level Review', *Annual Review of Organizational Psychology and Organizational Behavior* **3**(1), 55–81.

Tabrizi, B., Lam, E., Girard, K. & Irvin, V. (2019), 'Digital Technology is Not About Technology', *Harvard Business Review* pp. 2–7.
URL: *https://hbr.org/2019/03/digital-transformation-is-not-about-technology*

Thagard, P. (2017), Cognitive Science, *in* R. Frodeman, ed., 'The Oxford Handbook of Interdisciplinarity', 2 edn, Vol. 1, Oxford University Press, chapter 14.

Thau, S., Pitesa, M. & Pillutla, M. (2014), 'Experiments in Organizational Behavior', *Laboratory Experiments in the Social Sciences: Second Edition* pp. 433–447.

Tononi, G. (2004), 'An information integration theory of consciousness', *BMC Neuroscience* **5**(1), 42.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016), 'Integrated information theory: from consciousness to its physical substrate.', *Nature reviews. Neuroscience* **17**(7), 450–61.

Valorinta, M. (2009), 'Information technology and mindfulness in organizations', *Industrial and Corporate Change* **18**(5), 963–997.

Voegelin, E. (2000), *Order and History, Plato and Aristotle: Volume III*, University of Missouri Press, Columbia and London.

von Krogh, G. (2018), 'Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing', *Academy of Management Discoveries* **4**(4), 404–409.
URL: *http://journals.aom.org/doi/10.5465/amd.2018.0084*

Walsh, J. P. (1995), 'Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane', *Organization Science* **6**(3), 280–321.

Walsh, J. & Ungson, G. R. (1991), 'Organizational Memory', *Academy of Management Review* **16**(1), 57–91.

Weick, K. E. (1976), 'Educational Organizations as Loosely Coupled Systems', *Administrative Science Quarterly* **21**(1), 1.

Weick, K. E. (1989), 'Theory Construction as Disciplined Imagination', *Academy of Management Review* **14**(4), 516–531.

Weick, K. E. & Roberts, K. H. (1993), 'Collective Mind in Organizations: Heedful Interrelating on Flight Decks', *Administrative Science Quarterly* **38**(3), 357.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2005), 'Organizing and the Process of Sensemaking', *Organization Science* **16**(4), 409–421.
URL: *http://pubsonline.informs.org/doi/abs/10.1287/orsc.1050.0133*

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2008), 'Organizing for High Reliability: Process of Collective Mindfulness', *Crisis Management* **3**, 31–66.

Whetten, D. A., Felin, T. & King, B. G. (2009), 'The practice of theory borrowing in organizational studies: Current issues and future directions'.

Chapter 2

How swarm size during evolution impacts the behavior, generalizability, and brain complexity of animats performing a spatial navigation task[1]

[1] The idea for this paper is based on my master thesis submitted to the Institute for Intelligent Systems at the University of Magdeburg, May 2017.

Summary

While it is relatively easy to imitate and evolve natural swarm behavior in simulations, less is known about the social characteristics of simulated, evolved swarms, such as the optimal (evolutionary) group size, why individuals in a swarm perform certain actions, and how behavior would change in swarms of different sizes. To address these questions, we used a genetic algorithm to evolve animats equipped with Markov Brains in a spatial navigation task that facilitates swarm behavior. The animats' goal was to frequently cross between two rooms without colliding with other animats. Animats were evolved in swarms of various sizes. We then evaluated the task performance and social behavior of the final generation from each evolution when placed with swarms of different sizes in order to evaluate their generalizability across conditions. According to our experiments, we find that swarm size during evolution matters: animats evolved in a balanced swarm developed more flexible behavior, higher fitness across conditions, and, in addition, higher brain complexity.

Keywords: Artificial evolution, Multi-agent systems, Markov brains, swarm intelligence

Co-Authors:	Larissa Albantakis, Sanaz Mostaghim (see the Appendix for the declaration of the individual contribution).
Current-Status:	*Published [peer-reviewed]*, see: Fischer D, Mostaghim S, Albantakis L. How swarm size during evolution impacts the behavior, generalizability, and brain complexity of animats performing a spatial navigation task. GECCO 2018. 2018; doi:10.1145/3205455.3205646.
Permission:	Reprinted by permission from: ACM, 2018. License: See section 2, ACM Copyright Policy (Version 9 Revised 1/12/16).
Paper Presentation:	ACM GECCO, July 15–19, 2018, Kyoto, Japan.
Acknowledgements:	I gratefully acknowledge the informal support of Arend Hintze and Clifford Bohm from the Michigan State University. Further, I gratefully acknowledge the valuable feedback of the two anonymous reviewers, while proposing for the ACM GECCO'18 conference.

2.1 Introduction

When watching swarms in real life people often assume a global intelligence behind the swarm behavior, e.g., a flying mock of birds may seem to behave like a single organism (Garnier et al. 2007). However, we now know that individuals in a swarm often act based on local rules to achieve global goals (Reynolds 1987). This principle underlies the development of dedicated algorithms to solve single and multi-objective optimization problems, like Particle Swarm Optimization (PSO) or Ant Colony Optimization (ACO). Again, from the outside perspective, the abstract and virtual organisms seem to have swarm behavior, but their most basic rules are predefined by the optimization algorithms, e.g., in ACO all actions are predefined by the algorithm and its parameters (Dorigo et al. 1996). The observed complexity of swarm intelligence is then a result of the optimization process through interaction with the environment (Ilie & Badica 2013). Using machine learning approaches, it is also possible to evolve such swarm behavior without the need of a predefined algorithm that controls the organism, which means that the hard-wired decision rules of the earlier mentioned algorithms are now replaced by unsupervised learning techniques (Stanley et al. 2005, Olson et al. 2012, König et al. 2009).

Being able to evolve swarm behavior brings up new questions, e.g., about the effects of swarm size on evolution and how swarm size during evolution influences the organisms' decision rules. While in the scope of biology there are several studies on group size effects in swarms and optimal group size for different species (Pacala et al. 1996, Brown 1982), it is hardly feasible to conduct studies spanning evolutionary time-scales. Here, computational approaches using Evolutionary Algorithms (EAs) resembling evolution in nature provide new tools to address these questions. However, since the cognitive decision rules of adaptive artificial organisms are not hard-wired, but evolved over time they are readily observable but often hard to interpret.

In this study, we want to advance on these open questions by evolving animats[2] with swarm behavior and analyzing their internal 'brain' states and decisions. We hypothesized that swarm size could have a great impact on the evolution and social interactions of the animats, e.g., a fish in a swarm with 5 fish would behave differently than a fish in a swarm with 500 fish. For this reason, we investigated the effect of swarm size on the animats' evolution and, moreover, assessed the generalizability of their evolved behavior when performing in swarms of different sizes. In particular, we simulated and evolved groups of animats equipped with Markov Brains (MBs) (Hintze et al. 2017) in a novel spatial-navigation task environment that facilitated swarm behavior. As observed in previous work (Dorigo et al. 2004, Trianni et al. 2003), we found that the final environmental task fitness and its slope during evolution depended negatively on swarm size: task difficulty increased with swarm size as the 2-dimensional task environment became more crowded. In addition, however, the evolved animats showed significant differences in generalizability regarding their fitness when placed in different-sized swarms, which peaked for animats evolved in swarms of medium size. Interestingly, animats evolved in swarms of medium sizes also evolved more complex, integrated brain structures, which was evaluated by measuring the largest strongly connected network component in their MBs.

2.2 Related work

Adaptive animats equipped with the type of MBs used in this study were first introduced by Edlund et al. (2011). In several works, Olson et al. used these animats to investigated the evolution of swarm behavior in a predator-prey (co-)evolution environment (Olson et al. 2016, 2012). By contrast, swarm behavior in this study emerged directly as a result of the implemented selection rule (see below). The cognitive setup of animats used here to study group evolution and behavior most closely resembled the MBs described by Marstaller et al. (2013), which could move left or right and were evolved to solve a temporal-spatial integration task. The same type of task and animats were also used later by Albantakis et al. (2014), who evolved single animats in environments that required different degrees of context-dependent behavior and memory to investigate the evolution of integrated information (Oizumi

[2] An animat is an artificial animal with the ability to have specific motor reactions to sensory signals and to have internal states (Wilson 1985).

et al. 2014), a measure of brain complexity developed with the objective to capture the quality and quantity of consciousness in organisms.

A different approach to the evolution of artificial swarm behavior is the Neuroevolving Robotic Operatives (NERO) video game combined with real-time Neuroevolution of Augmenting Topologies (rtNEAT) (Stanley et al. 2005, Miikkulainen et al. 2012, Karpov et al. 2015). Similar to Olson et al. (2016), they also evolved swarm behavior in a predator-prey scenario, but with a learning technology based on Artificial Neural Networkss (ANNs). Miikkulainen et al. (2012) reviewed the work around neuroevolution and discussed future research topics in this field. They concluded that cooperative multi-agent systems are the next frontier of neuroevolution and that research in this field is still in an early stage.

Another alternative to animats equipped with MBs, could be Intelligent Distribution Agents (IDA) (Franklin et al. 1998) or, if adding the self-learning component, Learning IDA (LIDA) (Franklin & Patterson 2006, Franklin et al. 2012). The goal using this architecture was to develop a model of human cognition to investigate answers about the human brain and to apply the agents in real-life communications with humans.

Dorigo et al. (1996), who also developed the well-known algorithm Ant System (AS), implemented the evolution of self-organizing swarmbots as real robots and also investigated changing group size (Dorigo et al. 2004). They showed that it is easier for smaller groups to organize themselves as for larger groups. An earlier work considering different numbers of agents in the environment was conducted by Trianni et al. (2003), who also demonstrated that fitness decreases with increasing group size in a similar task of self-organization.

Apart from the above research on artificial systems, several studies investigated the subject of different swarm sizes in biological systems. Pacala et al. (1996) argue, on the example of ants foraging on food sources or nest maintenance, that a variation in swarm size implies that the organisms transfer different information and perform different tasks. It is mentioned that larger swarms can be more efficient than smaller ones, but very large swarm sizes can also be of disadvantage. In general, swarm behavior is the result of individual interaction with the environment combined with social interaction. Earlier, Brown (1982) presented work on the threshold of sociality, the willingness of an organism to join a swarm depending on the environmental qualities and swarm density. Here, optimal swarm size is expressed as a compromise between advantages gained by sharing costs and disadvantages arising from the faster loss of resources.

2.3 Methods

In order to design an environment in which the animats are able to evolve swarm behavior and allow for efficient analysis of their behavior and internal states, we have identified the following constraints to frame our model:

1. Animats must be able to co-exist (multiple organisms in one environment).
2. Respecting other animats (non-egoistic behavior) should help to gain higher fitness.
3. The task should be simple enough to be solved by animats with only a small number of sensors, motors and hidden nodes.

In this section, we describe the three main simulation components: (1) the animat design, (2) the 2-dimensional grid-based environment design, and (3) the EA's fitness function. The EA was configured using the Modular Agent Based Evolver (MABE) framework[3] (Clifford Bohm, Nitash C. G. 2017) for digital evolution. If not specified otherwise, we used MABE's default parameters throughout, which can be reviewed in the repository.

2.3.1 Animat design

Each animat used in this simulation contains a set of 2 sensors (one for walls, one to detect other animats), 2 motors, and 4 hidden memory nodes. Figure 2.1 shows a schematic illustration of the animats' architecture.

[3] See `https://github.com/Hintzelab/MABE/`.

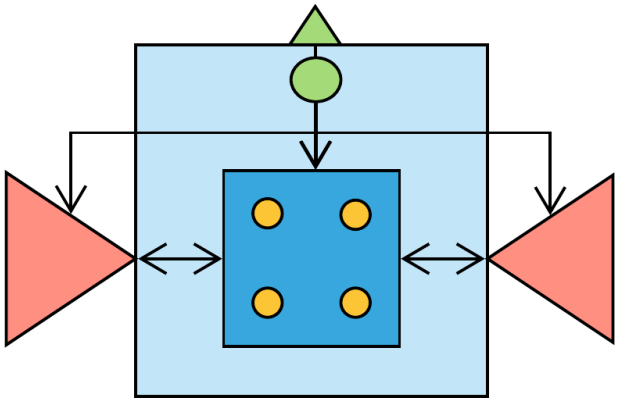


Figure 2.1: *Animat architecture. The green triangle marks the sensor for the wall, the green circle marks the sensor for other animats in the environment, the yellow circles mark hidden nodes and red triangles mark the motors. Sensors only connect to the hidden nodes and motors in a feedforward manner. Hidden elements and motors can feed back to all other nodes except sensors.*

The green triangle and circle mark the two sensors, red triangles indicate the motors, and yellow circles are the hidden nodes. Each animat has two kinds of sensors: one sensor detecting obstacles, here the walls (green triangle), and one sensor detecting other animats (green circle). Both types of sensors have a range of 1 unit directly in front of the animat. Animats are built with feedback motors. The motor elements can thus also act as memory just as the hidden nodes, which means that the current motor state at t_x can be causal for the future 'brain' state t_{x+1}[4]. All nodes in the network can have two states, `1` and `0`. A sensor switches to 1 if it detects an obstacle or animat, respectively. The movement model contains four possible states mapped by a 2-bit tuple $M = (m_l, m_r)$ where m_l and m_r model the left and right motor. $M = (1, 0)$ implies that only the left motor is active and therefore the animat turns left. The same holds for $M = (0, 1)$ in which the animat turns right. $M = (0, 0)$ indicates a static animat and $M = (1, 1)$ means that it moves forward.

Designing the animats with limited sensors and motors increases relative task difficulty, which has to be compensated by more complex internal states (Albantakis et al. 2014), and, for the same reason, should also facilitate the evolution of cooperation. This can also be observed in nature, e.g., for insects building their nests: As an individual their cognitive abilities are limited but as a swarm they are able to build complex structures (Theraulaz & Centre 1998). If the individual organisms were more intelligent, at some point swarming would not be essential for survival anymore.

Using the sensor data as the environment's representation, an animat has an internal representation stored in its artificial brain. There are several common types of such models: the brain could simply be a manually coded function, an ANN (Stanley et al. 2005), a simple finite state machine (König et al. 2009), or a MB (Edlund et al. 2011). In our case, the focus is on elucidating the animat's behavior while observing its internal and external states. Therefore, a simple and representable cognitive system was required. This is why we chose to implement MBs, which, moreover, emulate principles of neocortional function.

A MB is composed of a set of nodes with a finite set of states, which have temporal dependencies. The nodes' state-dependent update rules are implemented with the support of Hidden Markov Gates (HMGs), which indirectly connect the different nodes in the MB. Figure 2.2 shows an example architecture of a 4-node MB with one HMG. Each node can have a state (e.g., 1 or 0). A set of input nodes is connected to a HMG. Inside the HMG there could be a lookup table, or any other mechanism, transforming the inputs at t_0 into an output at t_1. The HMG's output is written to a set of output nodes, which could determine a motor state and/or be used as memory in the next time step. In this study, we exclusively used deterministic lookup tables to specify the HMGs' input-output functions.

[4] This is easily observable in the example wiring diagram below, Figure 2.11 on page 26.

The HMG's input-output functions and their inputs and outputs are encoded via a genome consisting of a string of integer values. At every generation, each locus in the genome has a probability of mutation; small sections of the genome also have a probability of being deleted or duplicated (Hintze et al. 2017, Edlund et al. 2011).[5]

Note that there is no active communication between the animats in this study. This means that the agents do not share information between each other, as, for instance, two organisms having a dialog. Animats can only sense whether another animat is directly in front of their position (or not).

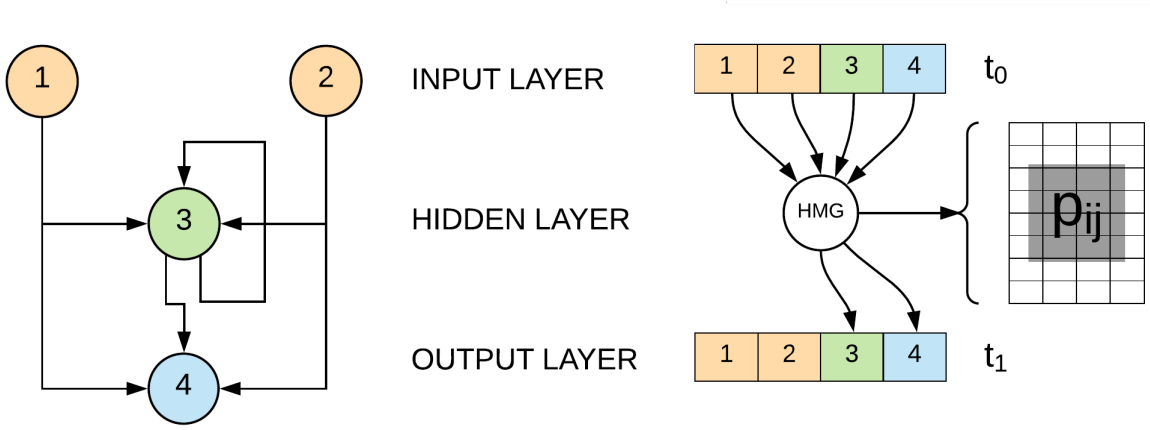


Figure 2.2: *A MB (Edlund et al. 2011) is composed of nodes, HMGs and their connections. The HMGs specify the mechanisms to transform a brain state at time t_0 to the future state at t_1, e.g., by fixed probabilities (indicated by p_{ij}). The effective brain connectivity between nodes (upper diagram) is derived from the nodes' hidden connections to and from the HMG.*

2.3.2 Grid-based environment and the challenge to move through the gate

In this work, we were interested in swarms and the evolution of swarm behavior, not just single animats as in previous studies (Edlund et al. 2011, Marstaller et al. 2013, Albantakis et al. 2014). Accordingly, multiple animats were placed in the environment simultaneously. Here, a swarm contained only clones, meaning each animat had the same genome and thus MB. Swarm size stayed constant across generations during each evolution. 5 different swarm sizes were tested. This made it possible to investigate dependencies between swarm size and the animats' (swarm) behavior. We distributed the swarm sizes uniformly up to a maximum of 72 animats, corresponding to all predefined starting positions in the task environment (see below):

1. $G_{1.00}$: 72 animats, the total number of available starting positions (constrained by the environment design).
2. $G_{0.75}$: 54 animats, 75 percent of the starting positions.
3. $G_{0.50}$: 36 animats, 50 percent of the starting positions.
4. $G_{0.25}$: 18 animats, 25 percent of the starting positions.
5. G_{single}: Only one[6] animat is placed in the environment.

The grid-based, 2-dimensional environment is designed to have 32×32 units (see Figure 2.3). At initiation, the animats were placed randomly without overlap on 72 predefined starting positions as marked in the figure by gray triangles. The environment was partitioned into two rooms connected only by a narrow gate, which the animats were supposed to cross frequently as part of their task. The design was inspired by the work of König et al. (2009).

The task environment was designed to pose a multi-objective problem. On the one hand, an animat received reward if it is able to travel through the gate. On the other hand, the animat got a penalty if it collided with other

[5] All parameters as in Clifford Bohm, Nitash C. G. (2017).

[6] Obviously, one animat cannot form a swarm. We included this condition here for comparison and treated it equivalently to simplify formulations.

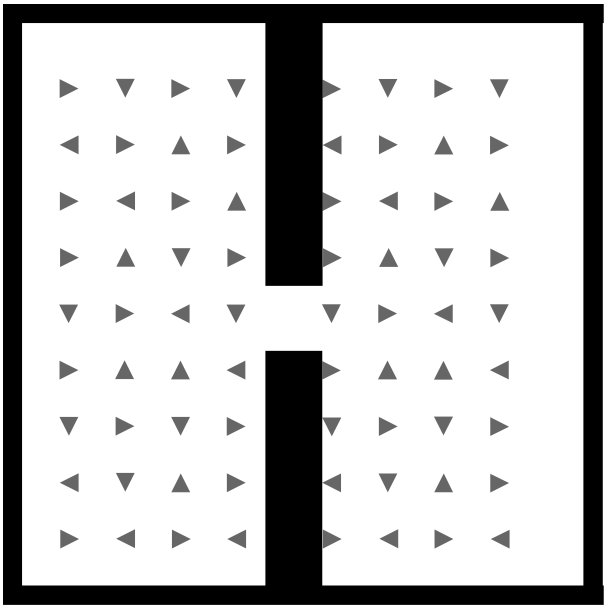


Figure 2.3: *Grid-based environment design. The environment contains two rooms, which are connected by a narrow gate. There are 72 fixed starting positions for the animats, onto which the animats are placed at random (except for $G_{1.00}$ where all positions are occupied), facing in random directions (up, down, left, right).*

animats (occupying the same location). Since such collisions between two animats were much more likely than crossing the gate, the collision penalty was set to a lower value than the reward for traveling through the gate, while still allowing for optimal behavior (0.075 compared to 1). If the penalty was chosen too low, all animats would learn to center around the gate area, colliding with each other all the time. If the penalty was chosen too high, all animats would adapt to not move at all (data not shown). Additionally, since it was not desirable for animats to crowd around the gate area, a timeout of 100 time steps was implemented until an animat can receive further rewards after crossing the gate once. Each trial had a total duration T of 500 time steps. This time-out period could be interpreted as simulating a requirement to make the way back to the organism's nest and also promotes the evolution of unified behavior.

2.3.3 Setup of the genetic algorithm

In the following we define the mathematical notation and equations for the fitness function. Table 2.1 lists all parameters and variables. In Equation (1) the fitness for a single animat in the environment is defined. Equation (2) specifies the overall fitness of the genome, or MB, which is also used for the selection process in the Genetic Algorithm (GA) of MABE. At each generation, we tested the swarm's genome 20 times in the environment ($|R|$= 20), with different random starting parameters[7] to obtain robust fitness values for each genome. In each of these 20 trials, we randomly picked a single animat out of the swarm[8] and averaged across their fitness values to obtain the overall fitness assigned to the genome.

$$f(a) = \sum_{t=0}^{T-1} \begin{cases} 1 & g(a,t,t+1) = 1 \text{ and } g(a,t-100,t) = 0 \\ 0 & \textit{otherwise} \end{cases} - \sum_{t=0}^{T} \begin{cases} 0.075 & c(x(a),y(a),t) > 1 \\ 0 & \textit{otherwise} \end{cases} \tag{2.1}$$

$$F(A,R) = \frac{\sum_{i=1}^{|R|} f(\text{randA}(A,R_i))}{|R|} \tag{2.2}$$

[7] Referring to the starting position and orientation.

[8] This was done in order to maintain the same sample size across conditions with different swarm sizes.

Table 2.1: *Definition of the mathematical notation for the fitness function*

a	Identifier of a single animat a, where $a \in \mathbb{N}$.
A	The set of all animats a in a trial, i.e. a swarm.
R	The set of all trials R_i an animat is tested in.
$f(a)$	The fitness of a single animat a.
$F(A, R)$	The average fitness of a genome across all trials R.
$\mathrm{randA}(A, R_i)$	Picks a random animat a from the swarm A depending on the trial R_i.
$g(a, t_a, t_b)$	Returns the count of gate-crossings between time t_a and time t_b for a single animat a.
$c(x, y, t)$	Returns the count of animats at a specific position (x, y) at time t.
t	A single time step t, where $t \in T$ and $t \in \mathbb{N}$.
T	Trial duration, i.e. the number of all time steps t in a trial.
$x(a), y(a)$	Returns the x and y position of animat a.

A single evolution experiment was run for 10,000 generations. At each generation a population of 100 genomes was evaluated, encoding the animats' MBs. After each generation, a set of 100 genomes was selected (with the possibility for duplicates) to enter the next generation based on their fitness values and the selection rules of the GA. These genomes were then mutated according to the probabilities specified above. Note that the genome population should not be confused with the swarm size: a swarm is a set of clones with identical genomes and thus MBs. The population of genomes corresponds to the pool for selection. For each of the five G_i conditions we ran 30 evolution experiments with different random seeds.

2.4 Results

To address our research questions, we performed a multi-level analysis to evaluate the evolved genomes resulting from the GA. First, we compared the average fitness evolution of the 30 evolution experiments per condition of the different swarm sizes G_i. Second, we investigated the movement patterns of all agents in a swarm to answer whether the animats evolved swarm behavior. Third, we evaluated the animats' generalizability across swarm sizes. An animat is generalizable if it performs at high fitness when tested with different swarm sizes as it was trained in. We also report qualitative observations about behavioral differences between the various swarm sizes G_i. Finally, we applied a simple graph-theoretical measure to the animats' MBs as a proxy for brain complexity.

2.4.1 Evolution of fitness

First, we analyzed the evolution of fitness across all test groups G_i. As it can be observed in Figure 2.4, task fitness is strongly dependent on swarm size: the smaller the swarm the steeper the curve and the higher the final evolved task fitness. This result shows that the task, in general, can be solved without any cooperation and actually becomes more difficult for larger swarm sizes since colliding with other animats results in penalty (ref. Section 2.3.2).

2.4.2 Observation of swarm behavior

Secondly, we tested if the animats developed swarm behavior or only independent movements. For this purpose, we generated heat-maps highlighting the movement patterns of the animats during their trial. Figure 2.5 shows the 5 heat-maps of the best genomes for each swarm size condition at the final generation in the GA. We also generated and inspected animations of the swarm's evolved behavior (final generation) for each of the 30 evolutions per condition. The most common movement patterns fit a wall-following strategy, as in the study by König et al. (2009). According to Pacala et al. (1996), such a strategy qualifies as swarm behavior as it is a result of social interactions and interaction with the environment. This would mean that also the interaction with the wall is part of the swarm-behavior.

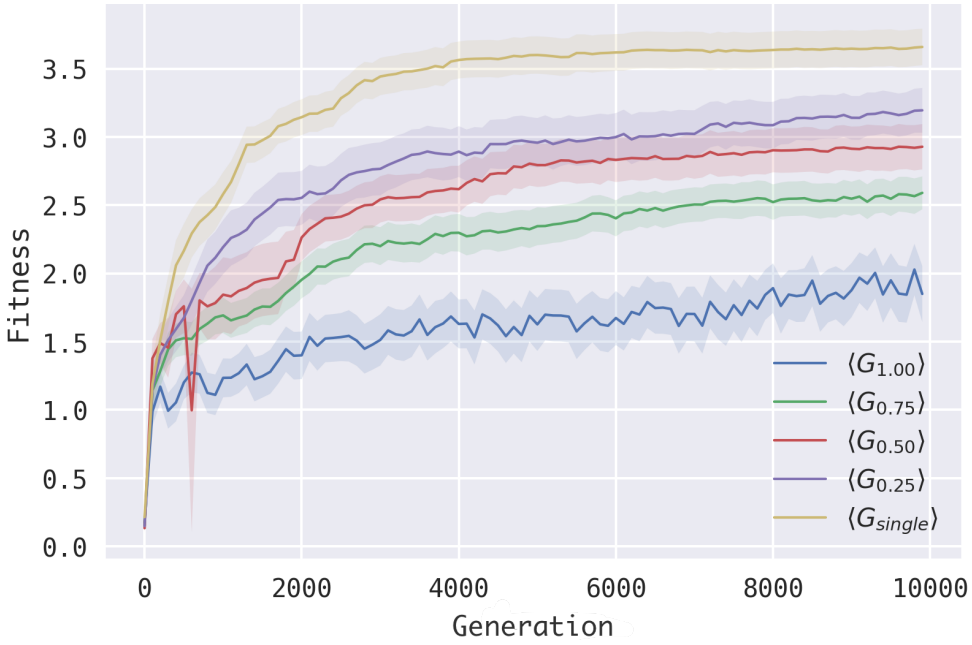


Figure 2.4: *Task fitness averaged across 30 evolutions of the five different configurations G_i. The overall fitness increase during evolution as well as the final fitness decrease with swarm size. The shaded area indicates the SEM.*

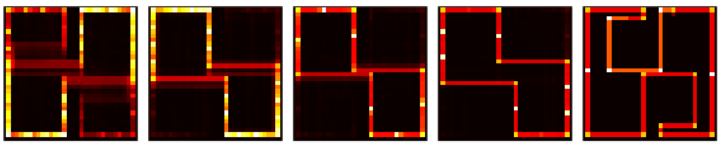


Figure 2.5: *Heatmaps of the best genome of all G_i. From left: $G_{1.00}$, $G_{0.75}$, $G_{0.50}$, $G_{0.25}$, G_{single}. Color indicates occupation density during the duration of one trial. Black areas were never visited, while red to white areas mark low to high density. Yellow/white cells and areas indicate spots where the animats turned or stalled frequently. This behavior was more common for animats evolved in large swarms.*

Most swarms with good fitness evolved such a wall-following strategy, but diversity in the movement patterns was also observed, particularly for animats with a lower final fitness. Nevertheless, only animats in G_{single} evolved high fitness using qualitatively different strategies. By distinguishing between (dark) red and yellow/white cells it is possible to observe whether the swarm is moving steadily or not. As the examples in Figure 2.5 show, big swarms moved slower along the walls, while smaller swarms only exhibited a few halting or turning points, particularly in the corners. Below, we first present results on the generalizability of the animats' behavior, followed by a quantitative analysis of the animats' sensory and motor states across conditions G_i (see below).

2.4.3 Generalizability of animats

To test the generalizability of the evolved animats, we tested the final generation of animats of all conditions G_i^{10k} in swarm sizes other than the one they evolved to, specifically, at $[100\%, 95\%, ..., 10\%, \text{single}]$ of the maximal swarm size of 72 individuals.[9] We then compared the robustness of their performance across swarm sizes to observe their generalizability (see Figure 2.6).

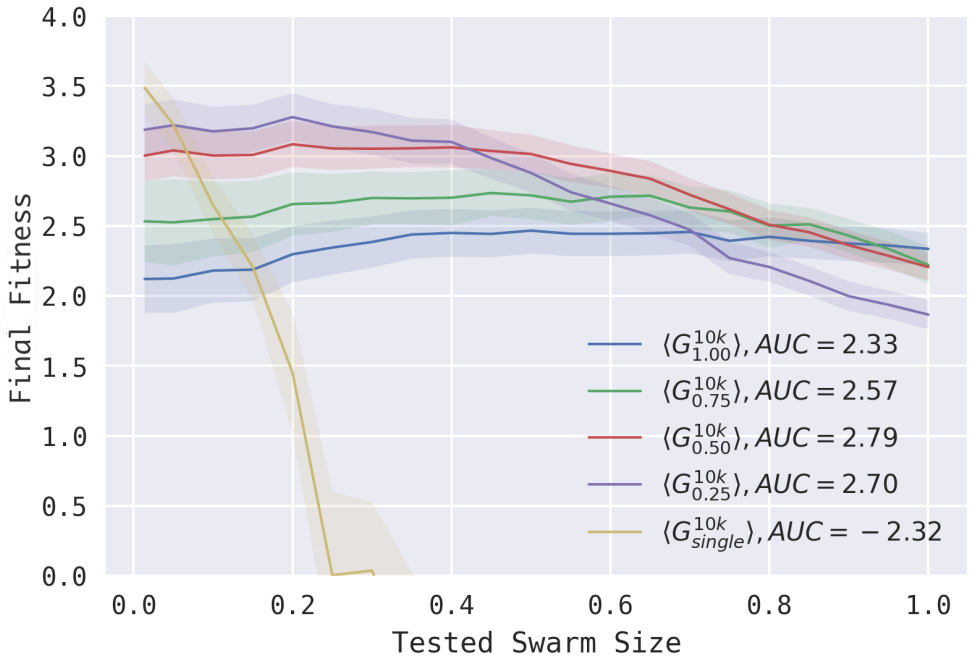


Figure 2.6: *Average performance of the final generation of each evolution experiment grouped by swarm size during evolution G_i^{10k} when tested at different trial swarm sizes* $[100\%, 95\%, ..., 10\%, \textit{single}]$.

While G_{single} animats failed to maintain their fitness within a swarm, all animats which were evolved in an actual swarm demonstrated a fair amount of generalizability. We quantified the fitness robustness of the different animat conditions G_i by calculating the Area Under the Curve (AUC), which is largest for the animats in $G_{0.50}$. Note also, that $G_{0.50}$ showed comparable fitness values to all other conditions except G_{single} in their original evolutionary swarm size. This suggests that adapting to intermediate swarm sizes may provide an advantage under changing environmental conditions, such as variation in swarm size due to rare environmental events. Our findings are also in line with Brown (1982) that too low and too high swarm density is negative for the overall swarm performance and respectively for the individual organism as well.

2.4.4 The cognitive processes of the animats

To identify regularities regarding the decision-rules in the respective swarm size conditions G_i we evaluated the animats' cognitive processes while performing the task. First, we evaluated the frequency of the animats' various

[9] This provided $3,000$ new trials: 5 swarm sizes $\times 30$ random evolutions $\times 20$ additional swarm sizes to test.

motor states (see Figure 2.7) of all final animats G_i^{10k} while performing the task in different swarm sizes (same data as in Figure 2.6). Here, variation across conditions indicates cognitive flexibility. What is more, these data also allowed us to differentiate whether being part of the swarm made the animats act in a certain way, or if, in turn, the swarm is merely the result of individual reactions. For G_{single} it should be obvious that individual animats were not influenced by the swarm and if seeming swarm behavior was observable it was not due to interactions with other animats. This is also supported by the data: G_{single} shows no variation in its motor responses across conditions. By contrast, animats evolved in swarms adapted their behavior dependent on the swarm size they were placed in. In particular condition $G_{0.50}^{10k}$, which demonstrated the greatest generalizability (see Figure 2.6), also had the most dynamic reaction to the different swarm sizes.

Figure 2.7: *Absolute count of movements, turns and no-movements over all G_i averaged over the different population sizes.*

Second, we measured how often specific sensory-motor state transitions could be observed (see Figure 2.8). Specifically, we recorded which actions at t_{x+1} followed a particular input at t_x. This corresponds to a complete external representation of the animats, i.e. their input-output behavior. Having two sensors and two motors, there are 2^4 possible external states. These were condensed to 4 bits in order to capture the following information: (1) The animat senses a wall, (2) the animat senses another animat, (3) the animat turns left or right, and (4) the animat moves forward (2 motors active). Note that, because of the nature of the task environment, instances such as sensing a wall and an animat at the same time or turning and moving forward at the same time are impossible and thus not considered, leaving 9 different state transitions to be evaluated. As an example, 0101 indicates that an animat sees another animat at t_x and moves forward at t_{x+i}.

Figure 2.8 shows the cumulative state transition count during a trial, averaged across the 30 evolutions for each G_i^{10k}. Since the first-order statistics of sensor and motor states depend on the size of the swarm during a trial, each animat evolved in a particular G_i was tested on each swarm size in the interval of $[100\%, 75\%, 50\%, 25\%, single]$ of the environment's maximal capacity (72 animats). This means that we counted all occurred transitions in 5 different trials of different swarm size. Moving forward without previously spotting another animat is by far the most frequent action in all conditions. It can also be observed that animats in G_{single} simply ignore other animats colliding with them despite the penalty (0101), while the others instead evolved to turn (0110) in such a situation.

Finally, animats in G_{single} seem to rely less on sensing the wall for guidance (1010), which could mean that they overall react less to their environment but rather use memory to solve the task.

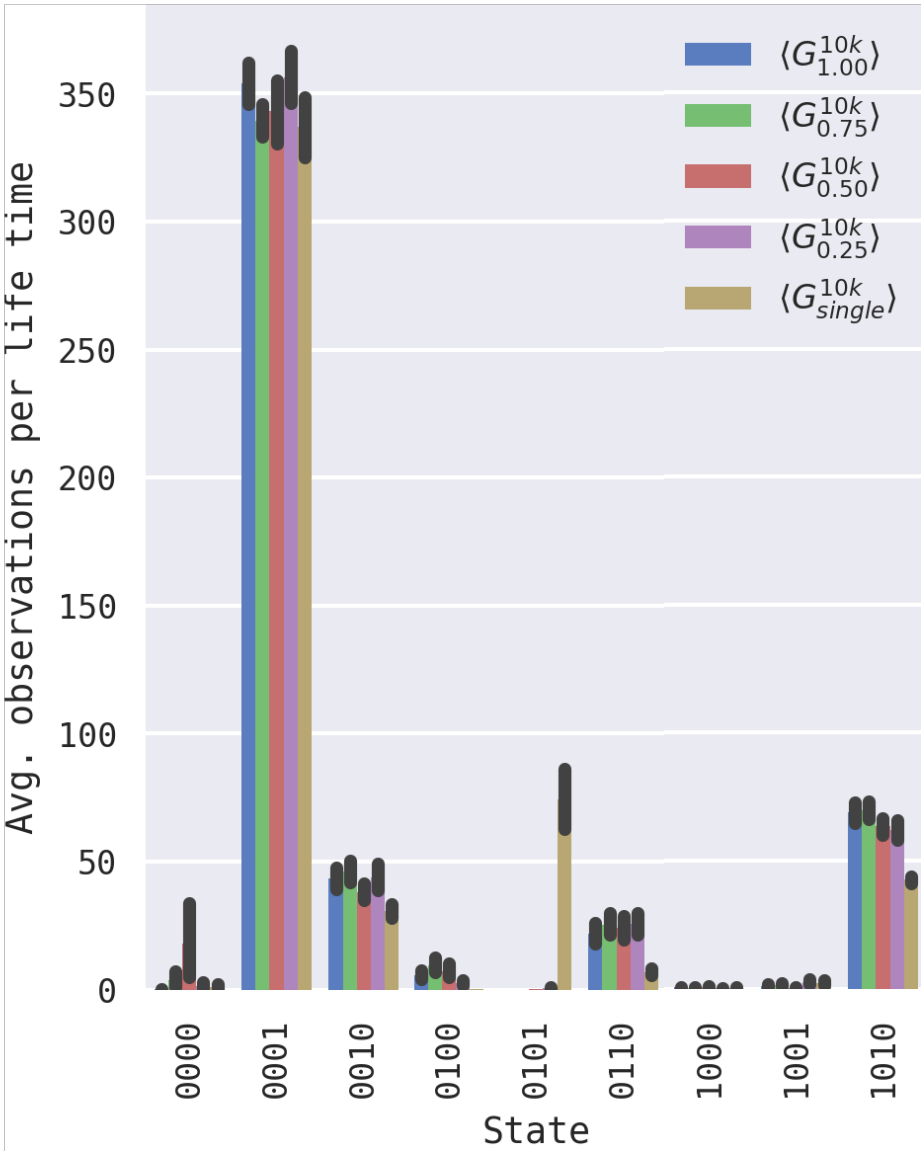


Figure 2.8: *The average number of times an animat enters a specific state transition, grouped by G_i^{10k}. The tuple is coded as (wall sensed, animat sensed, turn, move forward).*

To observe more detailed differences between the respective conditions than in Figure 2.8, we took one time step more into account and considered inputs at t_x, the reactions at t_{x+1}, the new inputs at t_{x+1} and the corresponding reactions at t_{x+2}. To visualize the data we generated a Transition Probability Matrix (TPM) (see Figure 2.9). For the sake of readability, we limited the labels in the plot and will thus describe them here. On the x-axis there are all inputs/reactions at t_x/t_{x+1}, on the y-axis there are all inputs/reactions at t_{x+1}/t_{x+2}. In the 9×9 matrix one tile visualizes the scaled probabilities per 3-time-step transition and G_i. Since we were more interested in differences across conditions than the absolute number of transitions, we scaled the bars according to their maximum and minimum values over all swarm conditions G_i, to better spot possible differences. Furthermore, values were averaged over all 30 different evolutions per condition, tested, as above, in five trials with different swarm sizes.

Animats in $G_{0.50}^{10k}$ stand out regarding their generalizability. Reviewing the TPM one can observe that such animats stayed static more often, especially when spotting another animat, and also stayed static in the following

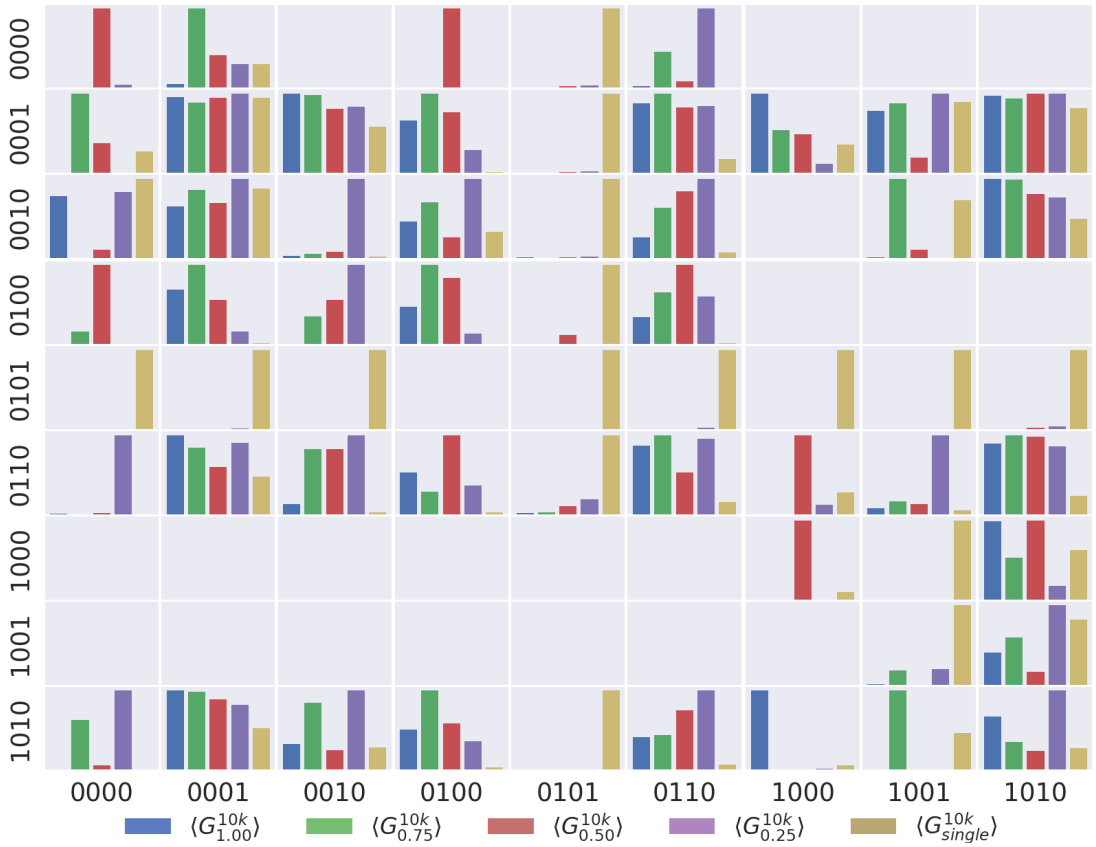


Figure 2.9: *External states of all G_i over three time steps. The axes are labeled with the state (see text for details). One tile shows how often the state transitions occur per G_i. The values of each single tile are scaled between 1 and 0, where 1 is the maximum probability to enter that transition and 0 simply the zero probability across conditions G_i.*

time step. This also supports our conclusion from Figure 2.7 that the behavior of animats in $G_{0.50}^{10k}$ was influenced most by sensing other animats. While it is observable that animats in G_{single}^{10k} always tried to move forward (spotting an animat or not), which had no negative effect in their original evolution environment, we observed that animats in $G_{1.00}^{10k}$ turned more often, even if they have no specific input.

2.4.5 Brain complexity

The results presented above indicate that $G_{0.50}^{10k}$ animats evolved the most complex behavior. Apart from the animats' externally observable input-output behavior, we also wanted to take their internal structure into account. The node connectivity in a MB can be modeled as a directed graph. As a simple graph theoretical measure of brain complexity, we thus used the Largest Strongly Connected Component (LSCC), which is also a simple measure of a graph's integration. As shown in Figure 2.10, animats acting alone or in large groups tend to evolve less complex brains even at high levels of fitness. Our assumption is that for G_{single} the environment was comparatively simple and rules to achieve high fitness were easier to find. Animats in $G_{1.00}$ could rely on sensing other animats with a high probability, which could serve as an orientation. $G_{0.50}$ evolved the most complex brain structures, which relates to our previous observations of the comparatively high behavioral complexity and generalizability of this group.

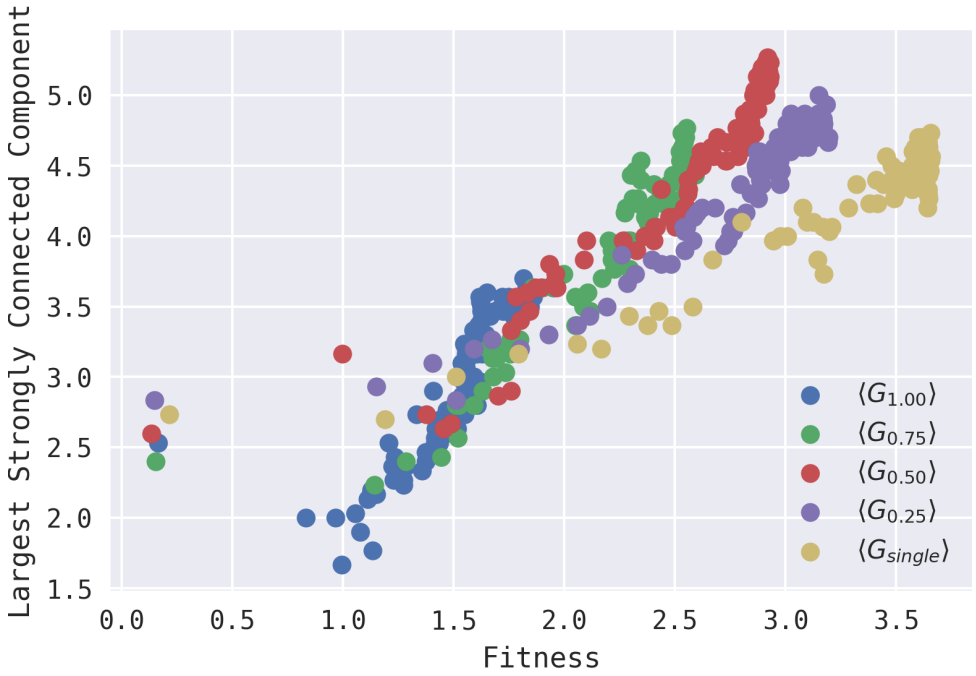


Figure 2.10: *Fitness plotted against the LSCC of the animat's brain, which we used as a proxy for brain complexity. One dot is the average LSCC over 30 experiments per generation evaluated at every* 100_{th} *generation).*

Figure 2.11 shows the wiring diagram of the best animat in $G_{0.50}^{10k}$ with an average task fitness of $F(A) = 3.8$. The animat has feed-forward sensors. All other nodes can feed back to each other, except to the sensors. This animat thus has the largest possible LSCC of 6. Feedback-loops are also an indicator for memory in a MB.

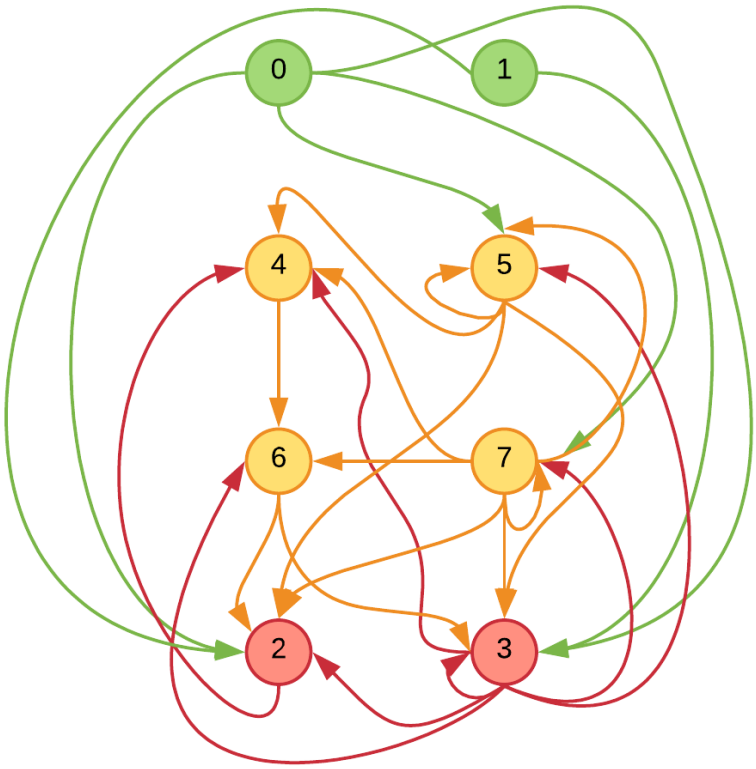


Figure 2.11: *Wiring Diagram of the best final genome in $G^{10k}_{0.50}$. The green circles mark sensor nodes, the yellow circles mark hidden nodes and the red circles mark motor nodes.*

2.5 Future work

The present results were obtained from one particular task environment. To build empirical strength, implementing more difficult and diverse tasks will be required, e.g., the predator-prey scenario used in earlier works by Olson et al. (2016) and Olson et al. (2016). Additionally, it is important to investigate variations in the animats' design to determine how the kind and number of sensors influence their behavior. In this work, animats were evolved and tested in an isolated manner. This means that a swarm only contained clones of one animat genome, which was necessary to make a first, specific evaluation. For future work, the effect of diverse swarms should be considered, which might increase the computational performance and make the simulated experiment more realistic. Finally, the development of more rigorous statistical analyses of the animats' external and internal state transitions is current work in progress.

2.6 Conclusion

Evaluating the detailed behavior and interactions of organisms in a simulated swarm is an open field of research. In this work, we addressed the effect of swarm size during evolution in a 2-dimensional spatial navigation task in a framework in which animats having Markov Brains were trained using a Genetic Algorithm. We, moreover, evaluated to what extent the resulting animats would be generalizable[10]. Furthermore, we focused on the evaluation of the animats' swarm behavior and its flexibility when faced with swarms of different sizes. We found that swarm size matters in the evolution of swarm behavior. Even if the task did not require cooperation, animats reacted to other animats non-egoistically in their decisions and formed swarm behavior. Our observation is that animats evolved in very large or very small swarms were less generalizable to other swarm sizes and showed less flexibility in their behavior. We assume that individuals in large swarms primarily acted to avoid collisions and the associated

[10] Testing animats in swarm sizes different from the one they evolved in.

penalty, while animats in small swarms had less incentive to develop proper reactions to encountering other animats. Overall, our results suggest that animats evolved at intermediate swarm sizes may have adaptive advantages due to their more generalizable and flexible behavior, which is also reflected in their higher relative brain complexity.

2.7 References

Albantakis, L., Hintze, A., Koch, C., Adami, C. & Tononi, G. (2014), 'Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity', *PLoS Computational Biology* **10**(12), e1003966. **URL:** *https://dx.plos.org/10.1371/journal.pcbi.1003966*

Brown, J. L. (1982), 'Optimal group size in territorial animals', *Journal of Theoretical Biology* **95**(4), 793–810.

Clifford Bohm, Nitash C. G., A. H. (2017), MABE (Modular Agent Based Evolver): A framework for digital evolution research, *in* 'Proceedings of the European Conference on Artificial Life', MIT Press, pp. 76–83.

Dorigo, M., Maniezzo, V. & Colorni, A. (1996), 'Ant system: Optimization by a colony of cooperating agents', *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **26**(1), 29–41.

Dorigo, M., Trianni, V., Şahin, E., Groß, R., Labella, T. H., Baldassarre, G., Nolfi, S., Deneubourg, J.-L., Mondada, F., Floreano, D. & Gambardella, L. M. (2004), 'Evolving Self-Organizing Behaviors for a Swarm-Bot', *Autonomous Robots* **17**(2/3), 223–245.

Edlund, J. A., Chaumont, N., Hintze, A., Koch, C., Tononi, G. & Adami, C. (2011), 'Integrated Information Increases with Fitness in the Evolution of Animats', *PLoS Computational Biology* **7**(10), e1002236.

Franklin, S., Kelemen, A. & McCauley, L. (1998), IDA: a cognitive agent architecture, *in* 'SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)', Vol. 3, IEEE, pp. 2646–2651.

Franklin, S. & Patterson, F. (2006), 'The LIDA architecture: Adding new modes of learning to an intelligent, autonomous, software agent', *Integrated Design and Process Technology* pp. 1–8.

Franklin, S., Strain, S., Snaider, J., McCall, R. & Faghihi, U. (2012), 'Global Workspace Theory, its LIDA model and the underlying neuroscience', *Biologically Inspired Cognitive Architectures* **1**, 32–43.

Garnier, S., Gautrais, J. & Theraulaz, G. (2007), 'The biological principles of swarm intelligence', *Swarm Intelligence* **1**(1), 3–31.

Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., Goldsby, H., Bohm, C. & Adami, C. (2017), 'Markov Brains: A Technical Introduction'. **URL:** *http://arxiv.org/abs/1709.05601*

Ilie, S. & Badica, C. (2013), 'Multi-agent approach to distributed ant colony optimization', *Science of Computer Programming* **78**(6), 762–774.

Karpov, I. V., Johnson, L. M. & Miikkulainen, R. (2015), Evaluating team behaviors constructed with human-guided machine learning, *in* '2015 IEEE Conference on Computational Intelligence and Games (CIG)', IEEE, pp. 292–298.

König, L., Mostaghim, S. & Schmeck, H. (2009), 'Decentralized evolution of robotic behavior using finite state machines', *International Journal of Intelligent Computing and Cybernetics* **2**(4), 695–723.

Marstaller, L., Hintze, A. & Adami, C. (2013), 'The Evolution of Representation in Simple Cognitive Networks', *Neural Computation* **25**(8), 2079–2107.

Miikkulainen, R., Feasley, E., Johnson, L., Karpov, I., Rajagopalan, P., Rawal, A. & Tansey, W. (2012), Multiagent Learning through Neuroevolution, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 7311 LNCS, pp. 24–46.

Oizumi, M., Albantakis, L. & Tononi, G. (2014), 'From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* **10**(5), 1–25.

Olson, R. S., Hintze, A., Dyer, F. C., Knoester, D. B. & Adami, C. (2012), 'Predator confusion is sufficient to evolve swarming behavior', *Journal of the Royal Society, Interface / the Royal Society* **10**, 20130305.

Olson, R. S., Knoester, D. B. & Adami, C. (2016), 'Evolution of Swarming Behavior is Shaped By How Predators Attack', *Artificial Life* **22**, 317–330.

Pacala, S. W., Gordon, D. M. & Godfray, H. C. J. (1996), 'Effects of social group size on information transfer and task allocation', *Evolutionary Ecology* **10**(2), 127–165. **URL:** *http://link.springer.com/10.1007/BF01241782*

Reynolds, C. W. (1987), Flocks, herds and schools: A distributed behavioral model, *in* 'Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques', SIGGRAPH '87, Association for Computing Machinery, New York, NY, USA, p. 25–34.
URL: *https://doi.org/10.1145/37401.37406*

Stanley, K. O., Cornelius, R., Miikkulainen, R., Silva, T. D. & Gold, A. (2005), 'Real-time Learning in the NERO Video Game', *Proceedings of the First Artificial Intelligence and Interactive Digital Entertainment Conference* **2003**, 2003–2004.

Theraulaz, G. & Centre, C. (1998), 'The Origin of Nest Complexity in Social Insects', **3**(6), 15–25.

Trianni, V., Groß, R., Labella, T. H., Şahin, E. & Dorigo, M. (2003), *Advances in Artificial Life*, Vol. 2801 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, Berlin, Heidelberg.

Wilson, S. W. (1985), 'Knowledge Growth in an Artificial Animal', *Proceedings of an International Conference on Genetic Algorithms and Their Appliations* pp. 16–23.

Chapter 3

How cognitive and environmental constraints influence the reliability of simulated animats in groups

Summary

Evolving in groups can either enhance or reduce an individual's task performance. Still, we know little about the factors underlying group performance, which may be reduced to three major dimensions: (a) the individual's ability to perform a task, (b) the dependency on environmental conditions, and (c) the perception of, and the reaction to, other group members. In our research, we investigated how these dimensions interrelate in simulated evolution experiments using adaptive agents equipped with Markov brains ("animats"). We evolved the animats to perform a spatial-navigation task under various evolutionary setups. The last generation of each evolution simulation was tested across modified conditions to evaluate and compare the animats' reliability when faced with change. Moreover, the complexity of the evolved Markov brains was assessed based on measures of information integration. We found that, under the right conditions, specialized animats could be as reliable as animats already evolved for the modified tasks, and that reliability across varying group sizes correlated with evolved fitness in most tested evolutionary setups. Our results moreover suggest that balancing the number of individuals in a group may lead to higher reliability but also lower individual performance. Besides, high brain complexity was associated with balanced group sizes and, thus, high reliability under limited sensory capacity. However, additional sensors allowed for even higher reliability across modified environments without a need for complex, integrated Markov brains. Despite complex dependencies between the individual, the group, and the environment, our computational approach provides a way to study reliability in group behavior under controlled conditions. In all, our study revealed that balancing the group size and individual cognitive abilities prevents over-specialization and can help to evolve better reliability under unknown environmental situations.

Keywords: Collective behavior, evolutionary algorithms, cognitive science, Markov brains.

Co-Authors:	Larissa Albantakis, Sanaz Mostaghim (see the Appendix for the declaration of the individual contribution).
Current-Status:	*Published [peer-reviewed]*, see: Fischer D, Mostaghim S, Albantakis L. How cognitive and environmental constraints influence the reliability of simulated animats in groups. Huk M, editor. PLoS One. 2020;15: e0228879. doi:10.1371/journal.pone.0228879
Permission:	Reprinted by permission from: PLOS, 2020. License: Creative Commons Attribution 4.0 International (CC BY 4.0).
Acknowledgements:	I gratefully acknowledge the support of Giulio Tononi and his department while staying at his research laboratory at the University of Madison-Wisconsin in Spring 2018. Further, I gratefully acknowledge the feedback of M. Huk and all anonymous reviewers while proposing this essay to the PLoS ONE journal.

3.1 Introduction

Intelligence is the ability to adapt to changes. According to this prevalent perspective, possessing general intelligence (Spearman 1904, Gardner 1987) not only enables one to perform a task correctly under already known conditions, but also to perform well under unexpected conditions. Further, in natural environments intelligent behavior is not only dependent on the (maybe limited) intelligence of the individual organism, but also involves interactions with the social and physical environment (Garnier et al. 2007, Pacala et al. 1996, Dorigo et al. 2004). The ability to adapt one's behavior to the behavior of other group members is necessary to act appropriately in case of unforeseen events, not only in the animal world but also in *high-reliability organizations*[1] (Weick et al. 2008, Weick & Roberts 1993, Oliver et al. 2017).[2]

While it seems intuitive that there is a triangular relationship between the individual, the group, and the environment (Fleck 1983), we discovered a lack of research on how individual behavior and group behavior are interrelated and depend on spatial attributes of the environment (Pinter-Wollman et al. 2018). Several studies have investigated intelligence and knowledge on the group level, and some have modeled groups of individuals as single agents, e.g., (Engel & Malone 2018, List & Pettit 2006, Walsh & Ungson 1991, Nonaka 2000, Tsoukas 1996). These studies have their origins in a variety of disciplines and have in common that they seek to elucidate the dynamics between group members. However, our understanding of how an individual actor in a group evolves intelligent behavior and reliability is still limited.

Here, we are particularly interested in how an individual's sensorimotor and memory capacity, the interaction between group members, and the environment constrain this evolution. To explore these factors in a controlled experimental setup, we used a simple evolution simulation, and we tested how specific cognitive and environmental limits influence the behavior, performance, and reliability of artificial organisms evolved in groups of various sizes.

Inspired and motivated by Pinter-Wollman et al. (2018), we investigated how the behavior and performance of evolved "animats"[3] varies in different task conditions, such as changes in the proportions of static objects, dynamic objects (moving group members), and individual sensorimotor and memory architecture. Using a simulation approach enabled us to manipulate and observe three dimensions which might influence evolved task performance and reliability: the group size (influencing the density of animats present in the environment), the animats' architecture (that is, the maximal number of available sensors, motors, and memory units), and the environmental design. In this study, we explicitly distinguish between the final task performance reached in the evolution environment ("evolved fitness" (EF)) and the post-evolutionary "task fitness" (TF), which measures the performance of the evolved animats under specific modified conditions (not encountered during evolution). High task fitness across many modified conditions indicates high reliability. High evolved fitness, but low reliability could then be interpreted as a form of narrow intelligence, while high evolved fitness and high reliability would point to more general intelligence.

We used a genetic algorithm to let the animats' behavior evolve under various evolutionary setups. Specifically, the animats were controlled by Markov Brains (MBs) (Hintze et al. 2018), which consisted of computational units whose functions and connectivity were determined by the animats' adaptive genome. The animats' task was to navigate through a two-dimensional world composed of two rooms without colliding with other group members (see Figure 3.1). Each animat could achieve a maximum score of 4 points within each trial, with a small penalty (-0.075 points) for each collision and a large reward ($+1.0$ points) for crossing gates between rooms. After an evolution of $10,000$ generations, we tested the final animats under modified task conditions modeled as: a variation in group size[4], the complexity of the static obstacles in the environment, and interaction rules between animats that affect task difficulty. The interaction rules include changes in the animats' ability to differentiate between static obstacles and other animats, the imposed collision penalty, and the possibility to inhabit the same location in the

1 E.g., aircraft carrier or nuclear power plants.
2 In the following, we use the term "reliability" to denote the ability of an organism to perform well even under slightly modified, unfamiliar circumstances.
3 Animats are simulated agents with cognitive abilities (Marstaller et al. 2012, Hintze et al. 2018).
4 The number of animats simultaneously present in the environment.

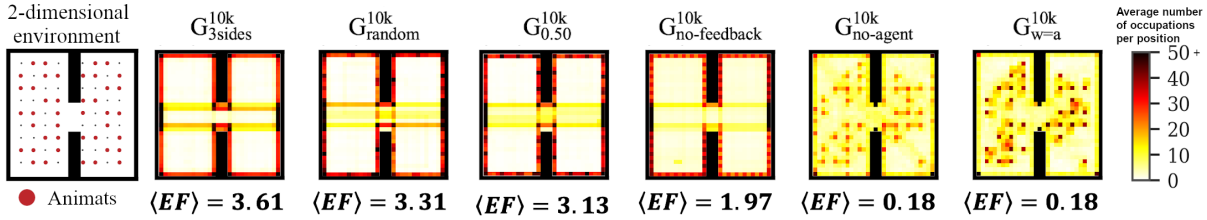


Figure 3.1: ***The average number of occupations per position in the final generations.*** *The first Panel on the left shows the two-dimensional environment, including two rooms with a total of* 72 *start positions (*32 *black dots [not occupied],* 32 *red dots [occupied]) for reference. In each trial, a subset of position is randomly selected as the animats' initial locations. The other six Panels show the average number of occupations per position as heat maps. The average is taken across time (*500 *time steps) and evolution simulations (*30 *per evolutionary setup). Red fields indicate high occupancy, and yellow fields indicate low occupancy in the corresponding position throughout the trial. Generally, well-performing animat groups evolve a wall-following strategy.* $\langle EF \rangle$ *indicates the mean evolved fitness of the final generation in the specific condition (see Results section for formal definition).*

environment. An animat was considered reliable if its task performance remained high across many variations of these test conditions.

A predecessor study focused on the influence of group size on the evolution of group fitness and reliability (Fischer et al. 2018), while the present work (1) extends the reliability experiments, (2) includes evolutionary setups with variations in the animats' architecture, and (3) elaborates the measurement of brain complexity by applying measures developed within the framework of the Integrated Information Theory (IIT) to the evolved MBs (Oizumi et al. 2014, Albantakis et al. 2014). There are two additional works which directly relate to our study: First, König et al. (2009) provided the original experimental setup. They designed a two-dimensional spatial-navigation task in which a swarm of robots has to learn to travel between two rooms. Second, Albantakis et al. (2014) showed how single animats evolve in a perceptual-categorization task environment with dynamic objects under various task difficulties. The primary motivation behind their work was to investigate the evolution of integrated information (Oizumi et al. 2014), which is an indicator for brain complexity, and its relation to task difficulty and memory capacity. Here, we discuss how the complexity of the MBs – evolved in the various experimental setups – is related to reliability as a prerequisite for general intelligence.

Overall, we found that, specialized animats can be reliable under the right conditions, that feedback from the motor units has an impact on performance and reliability, that animats benefit from passive interaction, and that more sensors enable reliability with simpler and less integrated brain structures.[5] Generally, our approach highlights the complexity of the dependencies between the three investigated dimensions: properties of the individual, group interaction, and environmental design. Even the simplified conditions of our simulation experiments make this complexity visible, and thus cautions against hasty generalizations, e.g., across different species or environments.

In the following, we will first present our results on the animats' task performance, reliability, behavior, and brain complexity across varying evolutionary setups. After that, we will discuss the findings in the broader scope of the literature and also how our work contributes to it. The last part of the work explains the methods and research design.

3.2 Results

We simulated the evolution of artificial organisms ("animats") with diverse cognitive architectures (number and type of available sensors, motors, and memory units) for 10,000 generations under various conditions. See Table 1 for an overview of all evolution simulations conducted.

[5] Which challenges the view that higher generalized intelligence is necessarily associated with more complex cognitive architectures.

All animats were evolved to travel between two rooms in a two-dimensional environment, which they shared with other animats of their same type ("clones" with the same genome), except in the "single" condition (see Figure 3.1(a) and Table 1). The evolutionary fitness selection occurs at the level of the genome (each generation consists of a population of 100 genomes) and is positively dependent on the average number of times that the corresponding animats ("phenotype") stepped through the gate ($+1.0$ points) between the two rooms. After a successful gate crossing, the same animat did not receive another reward for 100 time steps to avoid crowding at the gate. In addition, we imposed a small penalty each time they collided with other animats (-0.075 points, if not stated otherwise). Throughout, fitness values are displayed as absolute numbers with a maximum value of 4 points (corresponding to the maximal number of possible gate crossings without collisions). A detailed description of the task environments and the evolutionary algorithm is provided below in the Methods section.

In many evolutionary setups (see Table 3.1), high final fitness values ($EF > 3$, "evolved fitness") were reached. Figure 3.1(b) displays six different heatmaps visualizing several evolved movement patterns. It is observable that animat groups with reasonable Evolved Fitness (EF) converge towards a "swarm"-like wall-following behavior, which is determined by both, interactions with fellow animats and interactions with the environment (Pacala et al. 1996, Pinter-Wollman et al. 2018).

Once evolved, the best genome of each final generation was selected for post-evolutionary tests under modified conditions. Specifically, we modified the following three environmental factors: (1) the number of co-existing animats, (2) the complexity of static obstacles compared to the original two-dimensional environment (see Figure 3.1(a), and the Methods section for details on the environmental design), and (3) the interaction conditions between agents (see Table 2). For each test condition we assessed the Task Fitness (TF) achieved in the particular post-evolutionary test environment (to be distinguished from the animats' EF reached after 10,000 generations in its original evolutionary setup). In addition, we evaluated the animats' behavior and quantified their reliability (average task fitness across modified conditions) across varying group sizes in the original environment (R).

Finally, we quantified the complexity of the evolved MBs using two measures developed within the framework of IIT (Albantakis et al. 2014, Oizumi et al. 2014): the integrated information (ϕ_{Max}) and the corresponding number of concepts ($\#Concepts(\phi_{Max})$). The analysis was performed using `PyPhi`, the IIT Python toolbox (Mayner et al. 2018), using the standard settings according to (Oizumi et al. 2014). `PyPhi` takes the evolved MBs as an input in form of their Transition Probability Matrix (TPM). The TPM specifies how the states of the MB's computational units (e.g., motors and memory units) update, given the state of their inputs. In this study, all computational units are binary and deterministic (see Section 3.5.1). Briefly, ϕ quantifies how much of the information specified by all components of a system would be lost under a partition of the system. ϕ has been proposed as a measure of complexity, as it will be high for systems with many different components (functional differentiation) that are also highly integrated (Oizumi et al. 2014, Marshall et al. 2017). For a particular MB we identify the subset of computational units with the maximal amount of integrated information as ϕ_{Max}. For this subset, we also measure the number of components ("concepts") $\#Concepts(\phi_{Max})$. A "concept" in IIT is a subsystem that has a causal role within the system–a mechanism within the system. A concept causally constraints both, the past and future states of the system, and is irreducible to its parts. $\#Concepts(\phi_{Max})$ thus captures the number of internal functions performed by the subsystem with ϕ_{Max}. For details please refer to the original publication (Oizumi et al. 2014) and to Albantakis et al. (2014) for an application of these measures to evolved MBs. While there may be simpler, less computationally demanding options for evaluating the causal complexity of the evolved MBs (Marstaller et al. 2013, Hintze et al. 2018, Edlund et al. 2011), the chosen measures are fairly well established (Albantakis et al. 2014, Mayner et al. 2018, Marshall et al. 2018) and are theoretically motivated as part of the formal framework of the IIT (Oizumi et al. 2014).

We organized the presentation of our results into four sections categorized according to the evolutionary setups, as shown in Table 1 (varying "group size", "cognitive architecture", "interaction conditions", and "sensor configuration", respectively). Each section contains three figures displaying (1) the fitness evolution across generations and final evolved fitness values, (2) the task fitness, reliability, and behavioral features under modified

Table 3.1: *Definition of simulation conditions ("evolutionary setups"). Evolutionary setups are indicated by a label G_i, where the index i specifies the respective type of evolutionary setup.*

Label	G_i	Absolute Group Size[a]	Cognitive Architecture [b]	Interaction Condition[c]	Sensor Configuration	Result in Figures
Varying group size	***1.00***[d]	72	4 memory units 2 motors with feedback	Active penalty, blocking disabled	1 animat sensor, 1 wall sensor	3.2/3.3/3.4
	0.75	54				
	0.50	36				
	0.25	18				
	single	1				
	random	random				
Varying cognitive architecture	***bigbrain***	36	8 memory units	Active penalty, blocking disabled	1 animat sensor, 1 wall sensor	3.5/3.6/3.7
	smallbrain		2 memory units			
	no-feedback		4 memory units 2 motors without feedback			
Varying interaction conditions	***no-penalty***	36	4 memory units 2 motors with feedback	No penalty, blocking disabled	1 animat sensor,1 wall sensor	3.8/3.9/3.10
	blocked/no-penalty			No penalty, blocking enabled		
	blocked			Active penalty, blocking enabled		
Varying sensor configuration	***no-agent***	36	4 memory units 2 motors with feedback	Active penalty, blocking disabled	1 wall sensor	3.11/3.12/3.13
	3sides				3 animat sensors, 3 wall sensors	
	w=a				1 universal sensor	

[a] Absolute group size, 72 animats corresponds to 100% coverage of available starting slots.
[b] See Methods section for detailed architecture. Numbers indicate maximally available sensors, motors, or memory units, not the actually evolved number, which may be less.
[c] If penalty is active, animats receive penalty (-0.075 points) for colliding with other animats. If blocking is active, animats are not able to share the same position, otherwise they can occupy the same position, albeit with a penalty.
[d] Numeric indices correspond to relative group size: 1.00 corresponds to 100% coverage of available starting slots ($100\% \triangleq 72$ animats). The indicators 0.75, 0.50, and 0.25 correspond to 75%, 50% and 25% of available starting slots, respectively.

post-evolutionary test condition (see Table 2), and (3) a complexity analysis of the evolved MBs. Since the figures are redundant in their construction, we will briefly introduce their attributes:

Evolved fitness: Figures 3.2, 3.5, 3.8 and 3.11 show (a) the mean fitness $\langle F \rangle$ evolution across generations and (b) the distribution of evolved fitness values (EF) of the final generation across the $N = 30$ evolution simulations that we performed per evolutionary setup. The shaded areas in (a) visualize the standard error of the mean (SEM). The boxplots in (b) visualize the evolved fitness per condition G_i:

$$EF = F(A^i_{10,000}) \tag{3.1}$$

Where $A^i_{10,000}$ is the group of animats of the final generation of evolution simulation $i \in N$ and $F(A^i_{10,000})$ its fitness value (see Methods for more details on the fitness function).

Post-evolutionary tests: Figures 3.3, 3.6, 3.9 and 3.12 visualize the results of testing the final generation of animats across different group sizes ($GS = [1, 4, 7, ..., 65, 68, 72]$), Panel (a) in Figures 3.3/3.6/3.9/3.12, shows the mean task fitness $\langle TF \rangle$ of testing the animats under different group sizes in their original environment and under additional modifications of the interaction conditions between animats or the environment design, listed in Table 2. Note that the condition under which a group of animats evolved is indicated by their G_i label (see Table 1). $\langle TF \rangle$ is an average fitness across the $N = 30$ evolution simulations per experimental setup for a specific group size GS and (modified) condition M:

$$\langle (TF)^M_{GS} \rangle = \frac{\sum^N_{GS} F^M_G S(A^i_{10,000})}{N} \tag{3.2}$$

Table 3.2: *Overview of the eight environments in which reliability tests were performed. They differ in environmental conditions and in the complexity of the world design.*

Label	**Environmental Conditions**	**Environment (see Methods)**
Original	Active penalty[a], blocking disabled[b]	See Figure 3.16(a)
No penalty	No penalty, blocking disabled	See Figure 3.16(a)
Blocked	Active penalty, blocking enabled	
Blocked and no penalty	No penalty, blocking enabled	
Noisy Corners	Active penalty, blocking disabled	See Figure 3.16(b)
Small Gates		See Figure 3.16(c)
4 Rooms		See Figure 3.16(d)
4 messy Rooms		See Figure 3.16(e)

[a] If penalty is active, animats receive penalty (-0,075) when colliding into each other.

[b] If blocking is active, an animat cannot move onto the location of another animat.

Next, we quantified reliability for one test dimension, across modified group sizes in the "Original" test condition. We denote this specific measure of reliability as R, computed as:

$$R = \langle (TF)_{Original} \rangle^{GS} = \frac{\sum_g F_g(A^{10,000}_i)}{|GS|} \tag{3.3}$$

Note that in this case, the average is calculated across group sizes not evolution simulations as indicated by the subscript "GS", which stands for group size with $|GS| = 21$ (see above). Panel (b) shows the distribution of these reliability values (R) and their dependency on EF. Finally, Panel (c) shows how the animats' behavior depends on the relative group size in the "Original" test environment, evaluating the probability of an animat to stand still ("no movement"), turn, or move forward. Percentages are displayed in a scale from $0 - 100\%$.

MB Complexity analysis: Figures 3.4, 3.7, 3.10 and 3.13 show two types of metrics for MB complexity: (a) the distribution of integrated information (ϕ_{Max}) (Albantakis et al. 2014, Oizumi et al. 2014), and (b) the

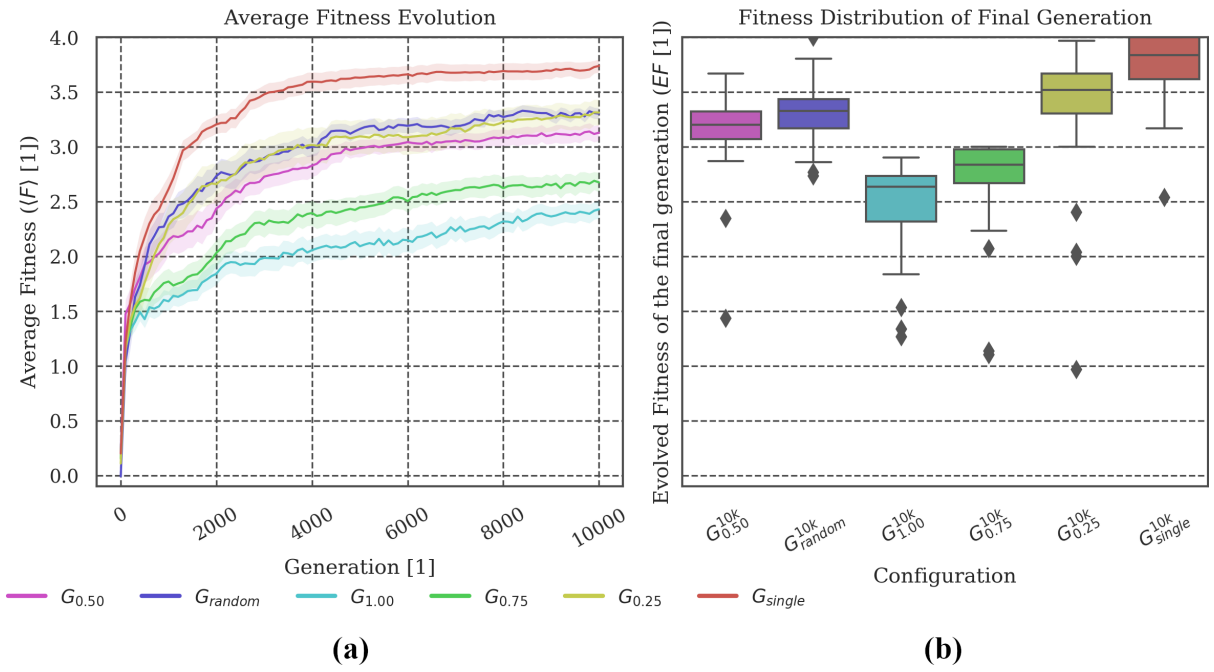


Figure 3.2: ***Fitness evolution and distribution of the final evolved fitness.*** *(a) G_{single} is the condition which evolves the highest fitness on average. Larger group sizes during evolution apparently impede the animats' fitness evolution and lead to lower final evolved fitness values. (b) The evolutionary setup with randomized group sizes at each generation (G_{random}) demonstrates similar properties as those setups with fixed, intermediate group sizes ($G_{0.25}$ and $G_{0.50}$).*

corresponding number of concepts ($\#Concepts(\phi_{Max})$) (Oizumi et al. 2014) per evolutionary setup. ϕ and $\#Concepts(\phi_{Max})$ are dimensionless quantities and therefore have no unit.

3.2.1 Varying group size: Evolution under specialized conditions can produce reliable agents

In a first set of experiments, we compared animats that evolved within groups of different, fixed sizes ($1 - 72$ animats), using the baseline animat and environment design in all cases, see Table 1: $G_{1.0-single}$. Preliminary results, including a comparison of the reliability R of evolution conditions $G_{1.0-single}$, were presented in (Fischer et al. 2018). As shown in Figure 3.2 (a) and reported in (Fischer et al. 2018), group size during evolution does impact the animats' ability to perform the gate crossing task (see Figure 3.1(a)), which impacts the final evolved fitness EF.

In our spatial-navigation task, animats in condition G_{single} (group size of 1 animat) frequently find an optimal solution within 10,000 generations. We assume that this is due to the decreased difficulty of the task in this condition since colliding is impossible, and walls (static obstacles) may still guide the animat towards the gate. Increasing the number of animats in the environment seems to make it more difficult to navigate. Animats have to develop not only the ability to cross the gate, but also to avoid collisions with other group members, which would cause a penalty (Fischer et al. 2018). Reliability R across group sizes was found to be high if the animats evolved in an environment where the density of animats was balanced ($G_{0.50}$ and $G_{0.25}$) (see Figure 3.3(a,b) and Fischer et al. (2018)).

In our study, we included an additional comparison setup (G_{random}), for which group size varied randomly during evolution. We hypothesized that animats evolved in this setup should achieve high reliability R in the post-evolutionary tests since variation in group size would already be part of their evolution. As shown in Figure 3.2(b), the final fitness values EF for G_{random} were comparable to those evolution setups with fixed, intermediate group sizes ($G_{0.50}$ and $G_{0.25}$) – though still significantly different ($p < .05$), see Section 3.7 for all statistical tests.

Table 3.3: *Absolute difference between the state transition probability P of $G_{0.50}$ and G_{random} ($P(G_{0.50}) - P(G_{random})$). The first digit ($S$) describes whether anything (wall or other animat) is sensed (1) or not sensed (0), and the second digit (M) describes whether the animat moved/turned (1) or did not move/turn (0). Most notably, G_{random} animats performed more movements even in the absence of sensor inputs than $G_{0.50}$ ("$01 \rightarrow 01$").*

SM		t+1			
		00	**01**	**10**	**11**
t	**00**	0.0000	-0.0074	0.0000	-0.0001
	01	-0.0079	0.0606[a]	0.0136	0.0088
	10	0.0005	0.0100	0.0063	0.0063
	11	-0.0001	0.0119	0.0031	0.0157

[a] Negative values indicate that the transition is more frequent in G_{random}, while positive values indicate the opposite.

As hypothesized, R was found to be highest for G_{random} (see Figure 3.3). Notably, however, animats that evolved under specialized conditions with intermediate group sizes ($G_{0.50}$ and $G_{0.25}$) reached R values comparable to animats that already encountered variable group sizes during evolution (G_{random}) (see Figure 3.3). $G_{0.50}$ and G_{random} show similar $\langle TF \rangle$ values in the original environment setting, particularly for larger group sizes ($> 50\%$ relative group size) (see Figure 3.3(a)). Nevertheless, G_{random} animats evolved to higher TF for smaller group sizes, leading to comparable but still significantly different average R values ($p < .05$) (see Figure 3.3(b)).

While R quantifies reliability across modified group sizes in the Original test condition, the other post-evolutionary tests (see Table 3.2) may reveal further differences between evolutionary setups. For example, *Blocked* (in which animats cannot overlap) suggests a difference in strategy between $G_{0.50}$, $G_{0.25}$, and G_{random} (see Figure 3.3(a)): $G_{0.50}$ and $G_{0.25}$ are more severely affected by this deviation from baseline settings in which animats can overlap, albeit under a penalty. While animats evolved in G_{random} also experienced large group sizes with a higher likelihood of a penalty during evolution, $G_{0.50}$ and $G_{0.25}$ animats consistently faced only intermediate probabilities of colliding with other animats, which may have led to less effective strategies for avoiding collisions. In addition to varying group sizes, we also tested the final generation of animats in four environments with different wall arrangements (see Figure 3.3(a), bottom row). $\langle TF \rangle$ decreased to similarly low levels in all conditions, but least for evolutionary setups with larger group sizes. Note also that G_{random} demonstrated relatively low $\langle TF \rangle$ under modified wall arrangements. Thus, high reliability across one dimension (here, modified group sizes as evaluated by R) does not necessarily transfer to other dimensions (e.g., modified wall arrangements).

In terms of their behavior (see Figure 3.3(c)), animats in G_{random} were less idle and showed fewer turns and more steps forward in comparison with animats in $G_{0.50}$, particularly for large group sizes. This suggests that the movement in G_{random} is more fluid overall (see also Table 3.3). By contrast, the specialized animats display larger differences in behavior across group sizes. Please refer to Fischer et al. (2018) for a more detailed discussion of behavioral differences across evolutionary setups with fixed group sizes $G_{1.0-single}$.

Figure 3.4 shows the distribution of ϕ_{Max} and $\#Concepts(\phi_{Max})$ (Oizumi et al. 2014, Albantakis et al. 2014) as a measure of the complexity of the evolved MBs across evolutionary setups with different group sizes $G_{single-1.0}$ and G_{random}. While the evolutionary setups with the highest R values (G_{random} and $G_{0.50}$) do show the highest average values of ϕ_{Max} and the largest number of concepts (internal mechanisms), differences between conditions generally do not reach statistical significance ($p >= .05$) due to the large variance in the complexity values (see Section 3.7). We assume that it would require more data (simulation experiments per evolutionary setup) to refine the mean of the intervals enough to verify the observed trend. In our predecessor study (Fischer et al. 2018), a correlation of high EF and reliability R with high brain complexity was found using a simplified measure of brain complexity based on anatomical connectivity only. The integrated information measures employed here are sensitive to the causal interactions within the MBs and thus also capture functional aspects in addition (Oizumi

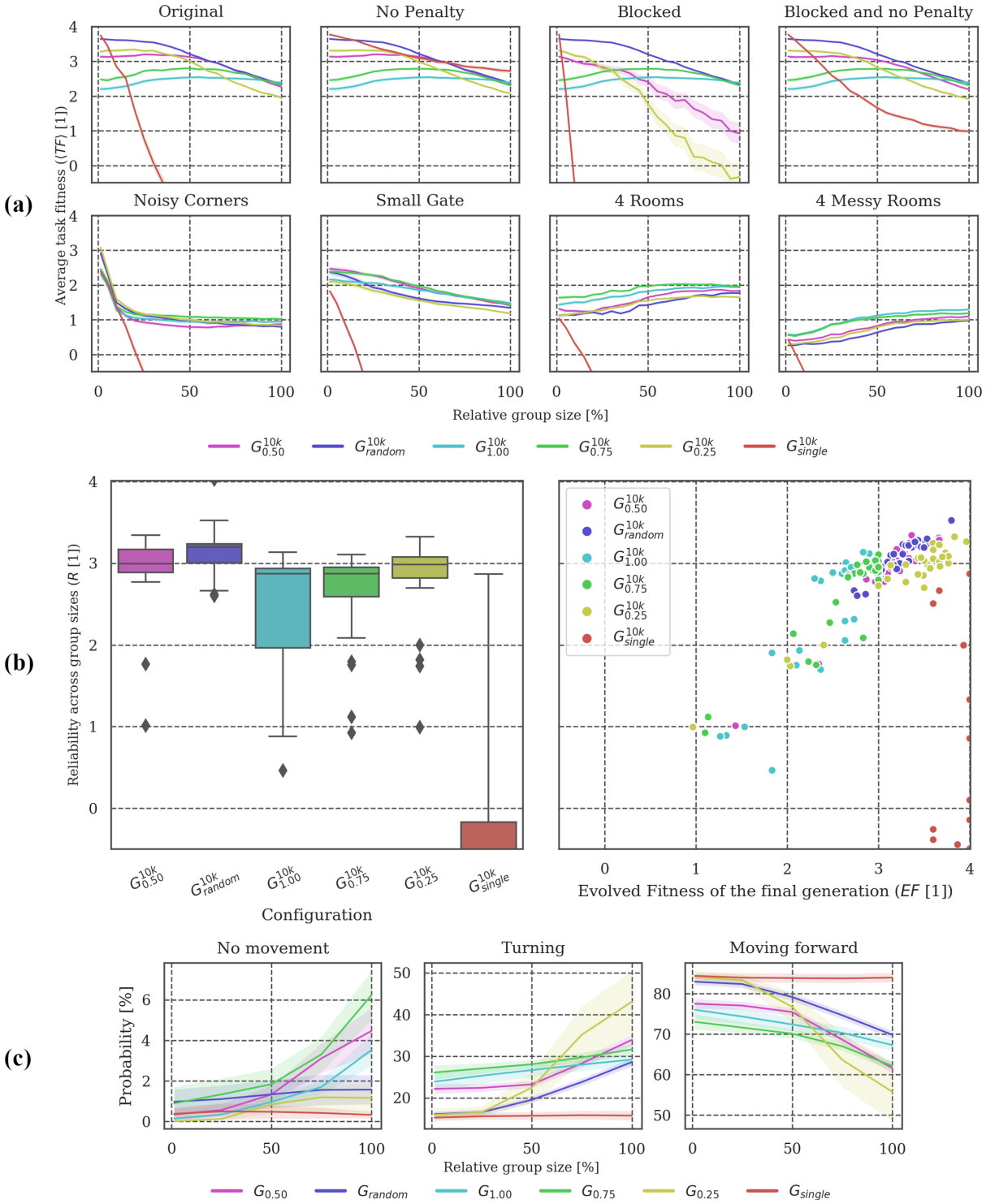


Figure 3.3: ***Post-evolutionary tests under modified conditions.*** *(a) Overall, only G_{single} failed to generalize across group sizes, presumably because animats that evolved without other group members did not develop strategies to avoid collisions (compare Original to No penalty test condition, where G_{single} performs well throughout). There is a large difference in the Blocked environment between G_{random}, $G_{0.25}$, and $G_{0.50}$, while in other environments their task fitness is comparable, pointing to somewhat different navigation strategies. (b) On average, G_{random} is the most reliable condition across varying group sizes, followed by $G_{0.50}$ and $G_{0.25}$. Except for G_{single}, EF correlates with R in all groups. (c) Note that $G_{0.50}$ and $G_{0.25}$ change their behavior more with increasing animat density compared to G_{random}.*

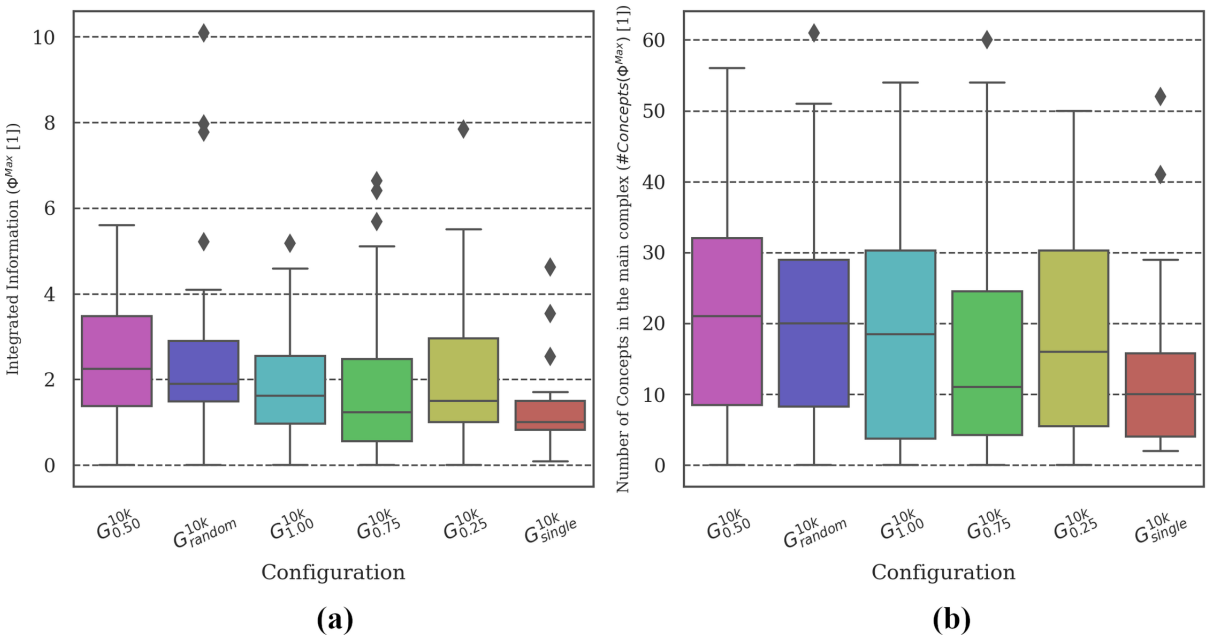


Figure 3.4: ***Distribution of brain complexity measures.*** *Differences in (a) ϕ_{Max} and (b) the corresponding number of concepts was found between the most (G_{random} and $G_{0.50}$) and the least (G_{single}) reliable setups. Due to the large variance in the data and the low sample size (*30 *simulations per evolutionary setup), differences in the mean between the remaining conditions did not reach statistical significance.*

et al. 2014, Albantakis et al. 2014). In the present data, significant pair-wise differences could be found between G_{single} and the most reliable setups (G_{random} and $G_{0.50}$). As explained above, the task environment experienced by animats in G_{single} is less demanding than for setups with larger group sizes. Our observations are thus in line with Albantakis et al. (2014), which demonstrated higher ϕ_{Max} and $\#Concepts(\phi_{Max})$ for animats evolved in more complex environments.

3.2.2 Varying cognitive architecture: Brain size and memory dependencies

In a second set of experiments, we used the same environmental setup as for $G_{0.50}$ in all tested conditions, but varied the number of available computational units in the animats' MBs. In the baseline design $G_{0.50}$, it is possible for the motor units to act as additional memory units (see Methods section). In one condition, $G_{no-feedback}$, the ability of the motor units to provide feedback was disabled, which reduced the absolute capacity for memory from six to four binary units. Moreover, we designed animats with similarly small memory capacity but with feedback motors as a reference group ($G_{smallbrain}$). Those animats had the original type of motors with the possibility of evolving feedback loops, but only two memory units instead of four. Finally, we included a condition with larger MBs with eight memory units and motor feedback ($G_{bigbrain}$).

We observed that evolved fitness EF and reliability R across group sizes in the original environment decreased for animats with fewer memory units (see Figures 3.5 and 3.6). However, while animats in $G_{smallbrain}$ still evolved to reasonably high fitness and reliability, $G_{no-feedback}$ was lacking in both. This observation indicates that motor feedback facilitates evolution in our task environment. One reason could be the fact that motor feedback allows the animats to utilize information about past movements directly (e.g., like the sensation of one's legs). One behavioral difference between $G_{no-feedback}$ and $G_{smallbrain}$ was the reduced movement in the animats of $G_{smallbrain}$ (see Figure 3.6(c)). Furthermore, the state transition analysis shows that the motor units of animats in $G_{smallbrain}$ tend to change their behavior more often, while animats in $G_{no-feedback}$ stay in the same state more often (see Table 3.4). Notably, $G_{no-feedback}$ and, particularly, $G_{smallbrain}$ performed better than $G_{0.50}$ in the *4 Rooms* and *4 Messy Rooms* test conditions (see Figure 3.6(a), bottom row).

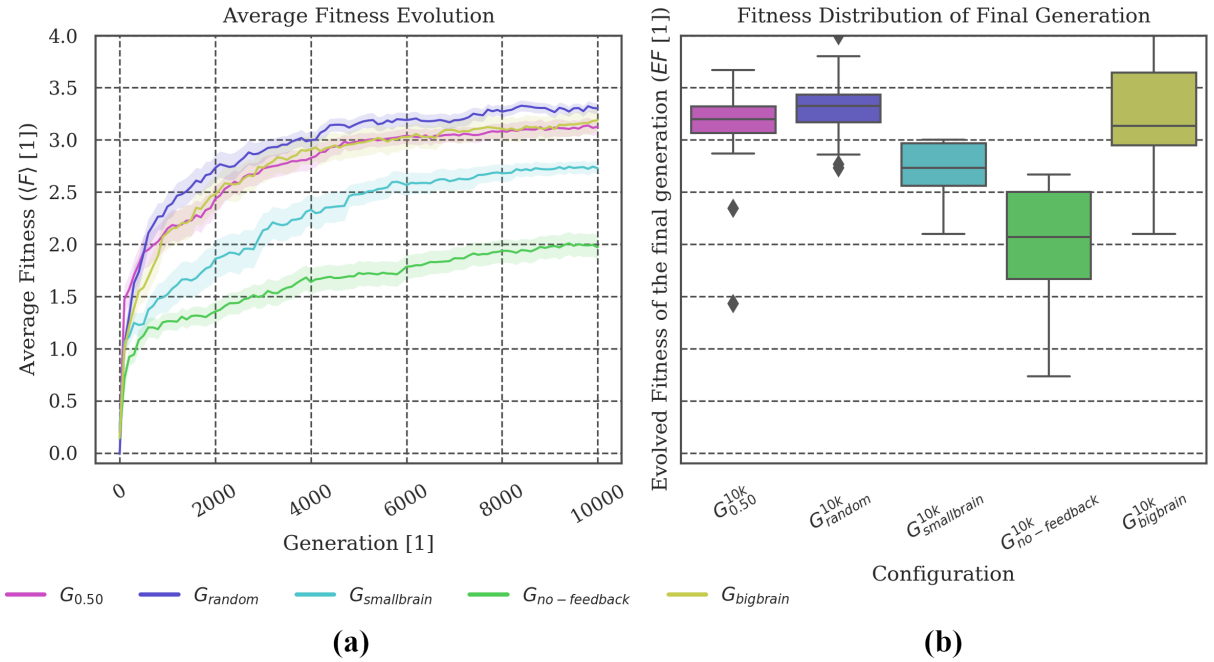


Figure 3.5: ***Fitness evolution and distribution of the final evolved fitness.*** *(a) Less capacity for memory and internal computations impairs fitness evolution. Despite their similar capacity for memory, $G_{smallbrain}$ evolved higher task fitness than $G_{no-feedback}$. (b) Ceiling outliers suggest that animats in $G_{no-feedback}$ are generally capable of performing as well as the average animat in $G_{smallbrain}$ but that this is less likely. The performance of $G_{bigbrain}$ is comparable to $G_{0.50}$ with more distributed outcomes.*

Table 3.4: *Absolute difference between the state transition probability of $G_{smallbrain}$ and $G_{no-feedback}$. The first digit describes whether anything (wall or other animat) is sensed (1) or not sensed (0) and the second digit describes whether the animat moved/turned (1) or did not move/turn (0). Most notably, animats in $G_{smallbrain}$ switched more often between sensing and moving than animats in $G_{no-feedback}$ ("$01 \rightarrow 10$", "$10 \rightarrow 01$", but "$11 \rightarrow 11$").*

SM		t+1			
		00	**01**	**10**	**11**
t	**00**	0.0000	0.0001	0.0000	0.0000
	01	0.0000	-0.0167[a]	0.0237	-0.0046
	10	0.0000	0.0194	0.0011	0.0029
	11	0.0001	-0.0004	-0.0015	-0.0241

[a] Negative values indicate that the transition is more frequent in $G_{no-feedback}$, while positive values indicate the opposite.

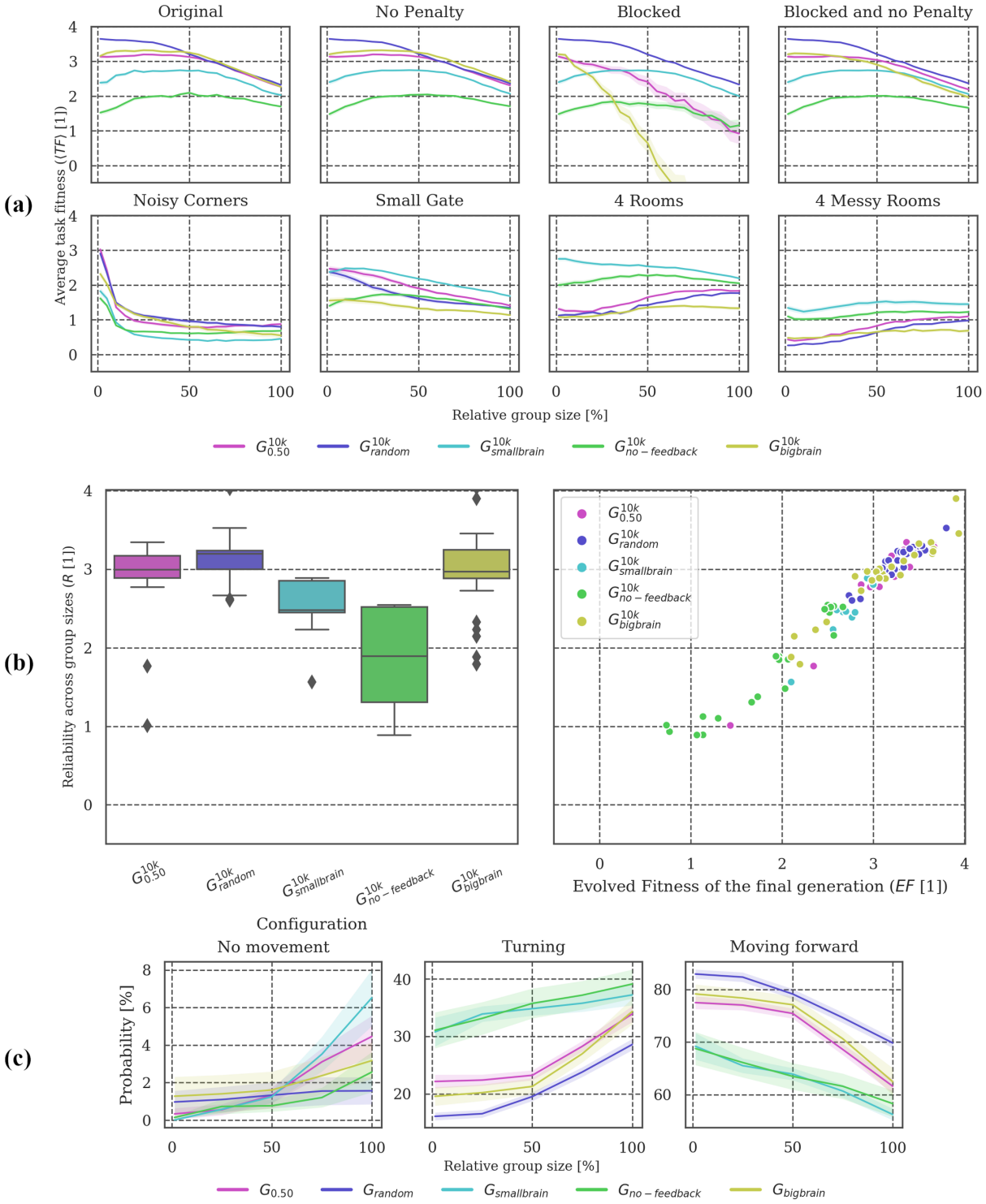


Figure 3.6: ***Post-evolutionary tests under modified conditions.*** *(a) $G_{smallbrain}$ shows higher $\langle TF \rangle$ than $G_{no-feedback}$ across group sizes. $G_{bigbrain}$ is overall comparable to the baseline condition $G_{0.50}$, but shows worse performance in the Blocked test condition and some of the modified environments for larger group sizes. (b) Reliability R correlates with EF for all setups. The lower R values of $G_{smallbrain}$ and $G_{no-feedback}$ compared to baseline can thus be explained by their already lower evolved fitness values. Note, however, that $G_{smallbrain}$ and $G_{no-feedback}$ perform better than $G_{0.50}$ across group sizes in the 4 (Messy) Rooms test conditions (see (a)). (c) For larger group sizes, $G_{smallbrain}$ remains static more often than $G_{no-feedback}$.*

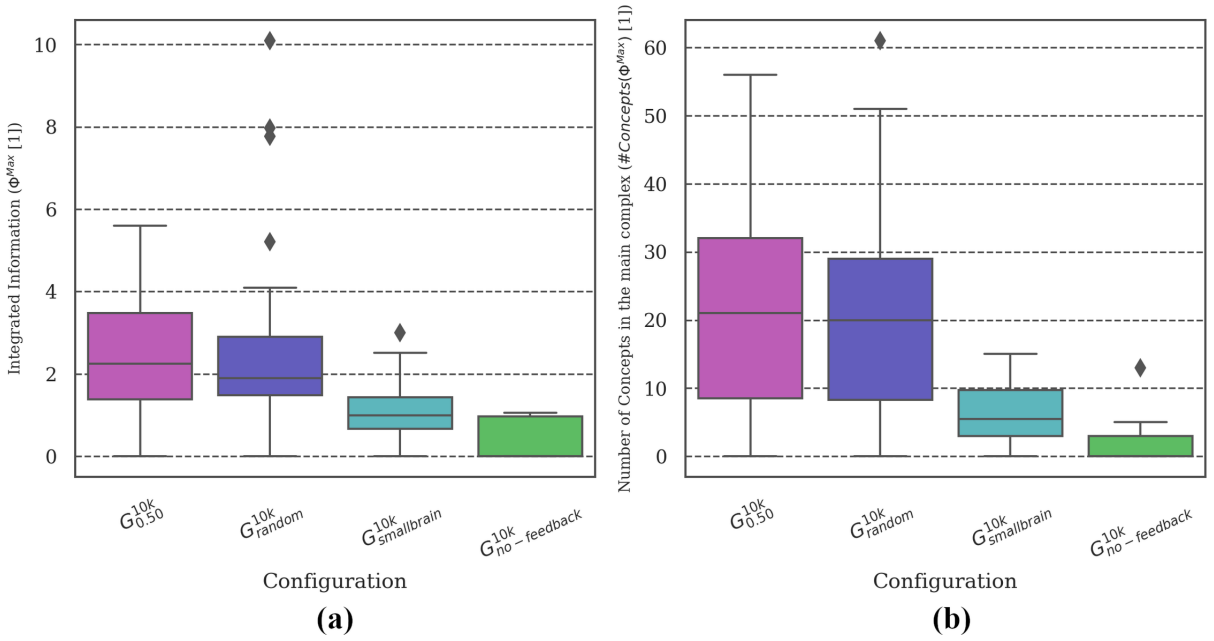


Figure 3.7: ***Distribution of brain complexity measures.*** *Compared to the baseline, the smaller MBs ($G_{smallbrain}$ and $G_{no-feedback}$) have lower ϕ_{Max} and fewer corresponding concepts. Animats in $G_{smallbrain}$ show higher ϕ_{Max} and have more corresponding concepts compared to $G_{no-feedback}$ animats, many of which have $\phi_{Max} = 0$. Due to computational reasons, the brain complexity of $G_{bigbrain}$ could not be calculated (see text).*

By contrast, more memory units ($G_{bigbrain}$) do not improve the fitness evolution or the TF in any of the tested conditions (see Figures 3.5 and 3.6). While $G_{bigbrain}$ achieves similar results compared to the baseline setup $G_{0.50}$, differences can be observed in the *Blocked* and *Small Gate* test conditions, as well as *4 (Messy) Rooms* for large group sizes (see Figure 3.6(a)). In principle more computational units should allow for better performance. However, the larger space of possible solutions may also impede fitness evolution (note the larger variance for $G_{bigbrain}$ compared to $G_{0.50}$ in Figure 3.5(b) and Figure 3.6(b)). Here, this trade-off may explain the similar mean $\langle EF \rangle$ and R values for $G_{0.50}$ and $G_{bigbrain}$.

Considering brain complexity, the evolutionary setups with smaller MBs ($G_{smallbrain}$ and $G_{no-feedback}$) have significantly lower ϕ_{Max} and fewer concepts than the baseline condition ($G_{0.50}$). Between those two conditions, $G_{smallbrain}$ shows significantly higher ϕ_{Max} and more concepts as compared to $G_{no-feedback}$ (see Figure 3.7). This correlates with the larger evolved fitness values of $G_{smallbrain}$ in Figure 3.5 and its associated higher reliability R in Figure 3.6. Note that calculating ϕ_{Max} and the corresponding number of concepts was not possible for $G_{bigbrain}$ since exhaustive evaluations across many systems and states are not currently feasible when using the `pyphi` software package to compute measures of integrated information theory for networks of that size (> 10 units) (Mayner et al. 2018).

3.2.3 Varying interaction conditions: Evolution of beneficial interaction

In our baseline configuration for the evolution simulations ($G_{0.50}$), individuals could occupy the same physical location but received penalties for colliding with other group members (see Methods section). We manipulated these features in the third set of simulations to evaluate how they influence both evolved fitness and reliability. Specifically, we considered three additional evolutionary setups: $G_{no-penalty}$, $G_{blocked}$, and $G_{blocked/no-penalty}$ (see Table 3.1 for a detailed description). G_{single}, G_{random}, and $G_{0.50}$ are also included in the figures for comparison.

Among the novel setups, only animats in $G_{blocked}$ were subject to the collision penalty during evolution. Not being able to share the same position (as in $G_{blocked}$) hardly influenced the EF, the mean task fitness $\langle TF \rangle$ across post-evolutionary conditions, or the behavior of the evolved animats compared to $G_{0.50}$ (see Figures 3.8 and 3.9).

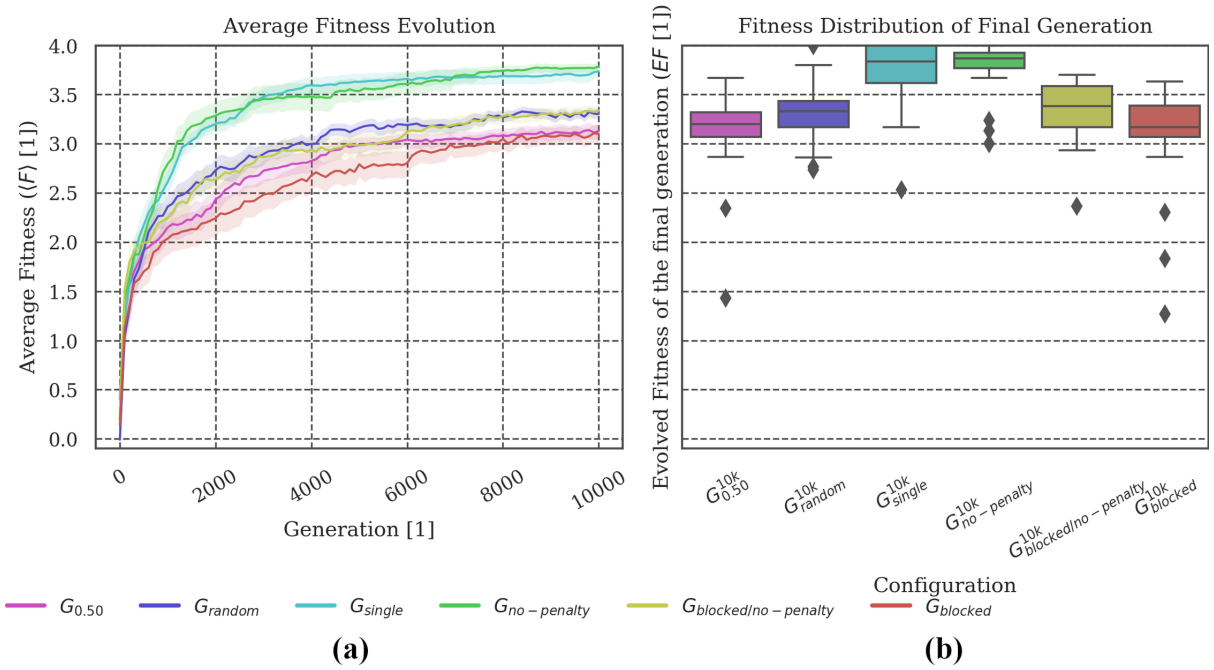


Figure 3.8: ***Fitness Evolution and distribution of the final evolved fitness.*** *The animats in conditions without a penalty ($G_{blocked/no-penalty}$ and $G_{no-penalty}$) evolved to relatively high fitness levels. In particular, $G_{no-penalty}$ evolved like G_{single}, which can be explained by the fact that animats in both of these conditions were not impacted at all by other animats. Similarly, $G_{blocked}$ seemed equivalent to the baseline setup $G_{0.50}$, while $G_{blocked/no-penalty}$ evolved to slightly higher fitness values, comparable to G_{random}.*

Likewise, $G_{no-penalty}$, where reacting to other animats had no direct effect on the fitness evolution, showed very similar EF, $\langle TF \rangle$, and behavior as G_{single}, with one exception: $\langle TF \rangle$ decreased with increasing group size in the *No Penalty* test condition for G_{single} but not for $G_{no-penalty}$ which had evolved with a group size of 36 animats, as in $G_{0.50}$ (see Figure 3.9(a)). Note that R in Figure 3.9(b) was evaluated in the *Original* task condition with penalty, as for all other simulations sets.

Considering the post-evolutionary tests in Figure 3.9(a), the top row shows $\langle TF \rangle$ across group sizes in the *Original* environment (with penalty) and under varying interaction conditions: *No Penalty*, *Blocked*, and both *Blocked and no Penalty* (from left to right). In the bottom row of Figure 3.9(a), animats are evaluated under the same interaction rules as they evolved in while only facing a modified environment (position of static obstacles).

In this context, it is noticeable that $G_{no-penalty}$ performed relatively poorly for larger group sizes when tested in *4 (Messy) Rooms* despite receiving no penalty for collisions. By contrast, in evolutionary setups with a collision penalty and/or blocking $\langle TF \rangle$ increased with group size in the *4 (Messy) Rooms* test conditions. The decline in $\langle TF \rangle$ of $G_{blocked/no-penalty}$ for larger group sizes under test conditions with a collision penalty (*Original* and *Blocked*) moreover, suggests that these animats did not avoid physical interactions with their group members. However, even $G_{blocked/no-penalty}$ animats had an advantage compared to $G_{no-penalty}$ in the *4 (Messy) Rooms* environment. Taken together, these observations let us assume, that any evolutionary pressure to *"pay attention"* to fellow animats (through blocking or a collision penalty) could lead to the evolution of interaction strategies with possible advantages under certain (modified) conditions (e.g., using other animats for orientation or guidance).

Considering the brain complexity of animats in $G_{blocked}$ and $G_{blocked/no-penalty}$, we can report similar values compared to $G_{0.50}$ (see Figure 3.10). In summary, whether animats received a penalty for crossing each other, or whether crossing was prohibited to start with, did not significantly affect their evolved fitness, reliability, behavior, or brain complexity. Likewise, the brain complexity measures and behavioral results for $G_{no-penalty}$ were comparable to those of G_{single}.

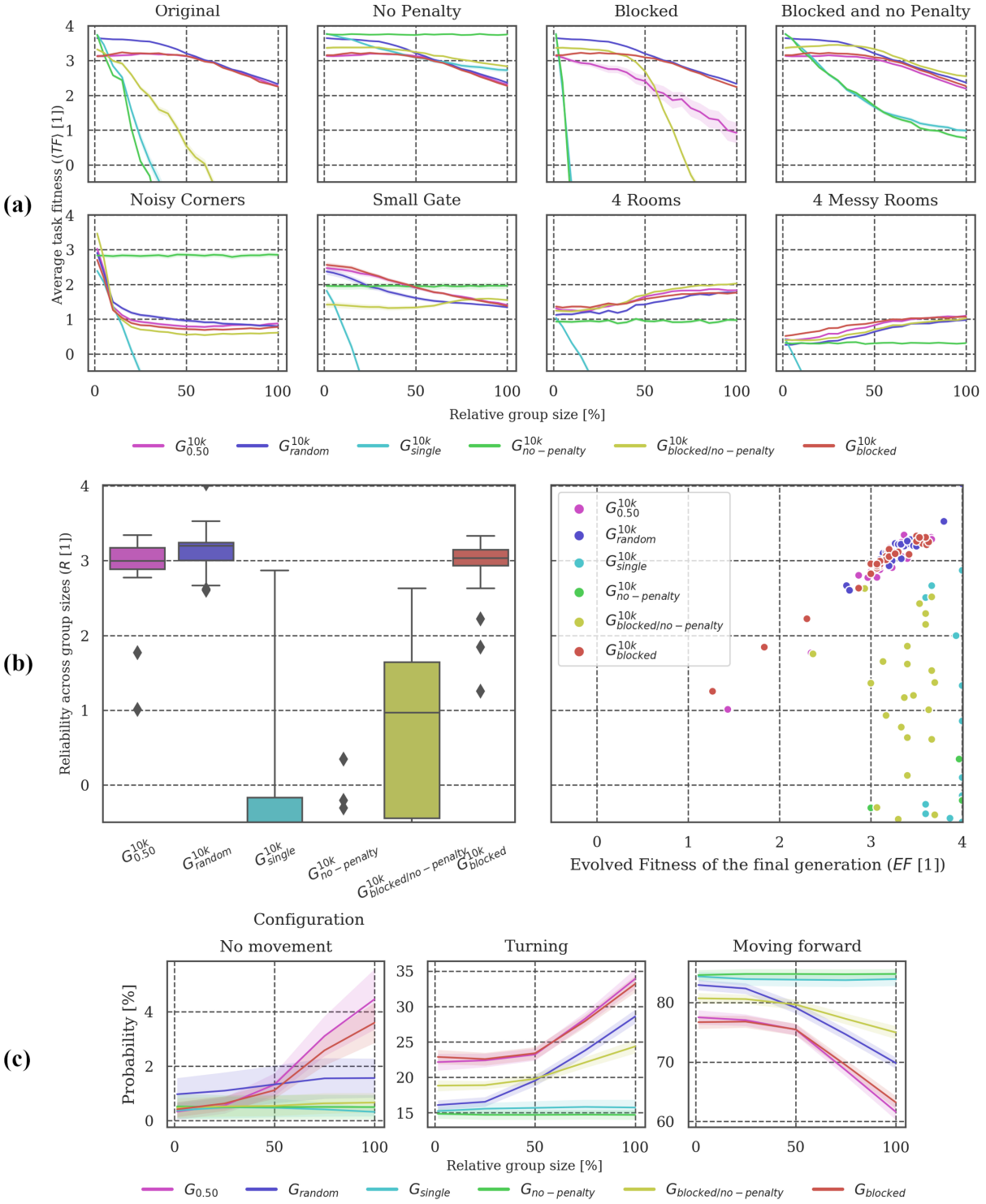


Figure 3.9: ***Post-evolutionary tests under modified conditions.*** *(a) There was a significant difference between conditions in which interactions with other agents played a role for fitness evolution ($G_{0.50}$, G_{random}, $G_{blocked}$, $G_{blocked/no-penalty}$) and those conditions in which it did not (G_{single} and $G_{no-penalty}$) (see text). (b) With a collision penalty imposed, $G_{no-penalty}$ showed similarly low reliability as G_{single}, whereas $G_{blocked}$ showed similarly high reliability as $G_{0.50}$. $G_{blocked/no-penalty}$ retained some reliability under collision penalty even though animats were evolved without it. (c) Similarities between $G_{0.50}$ and $G_{blocked}$, as well as G_{single} and $G_{no-penalty}$ were also reflected in the animats' behavior. The behavior of animats in $G_{blocked/no-penalty}$ was more reactive to changing group size than $G_{no-penalty}$.*

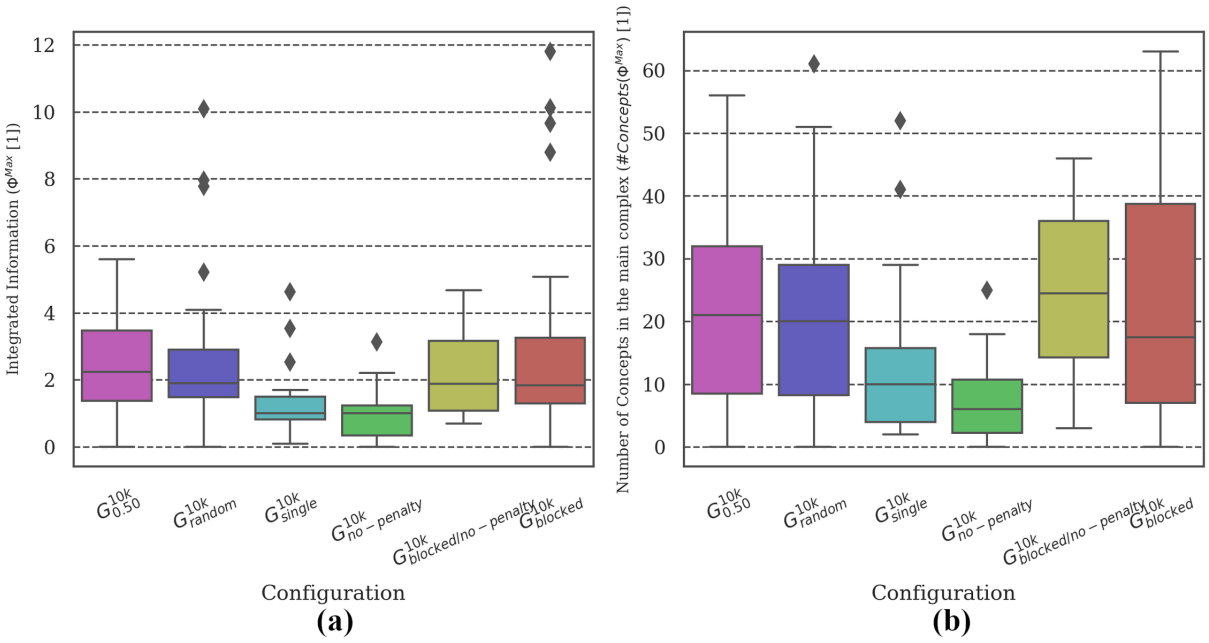


Figure 3.10: ***Distribution of brain complexity measures.*** *In evolutionary setups where crossing each other was not possible ($G_{blocked}$ and $G_{blocked/no-penalty}$), the brain complexity was comparable to the complexity of $G_{0.50}$. By contrast, animats in setups where the reaction to fellow animats had no reasonable effect on their performance (G_{single} and $G_{no-penalty}$) showed lower brain complexity. Still, there was high variance in the data of brain complexity.*

3.2.4 Varying sensor configuration: Sensory capacity influences reliability and brain complexity

We manipulated the animats' sensor configuration (see Table 3.1) in a final set of evolution simulations. In addition to the baseline architecture (front wall sensor and front agent sensor), we designed animats with sensors on three sides G_{3sides} (front, left and right wall and agent sensors), without an agent sensor $G_{no-agent}$ (one front wall sensor only) and with one universal sensor $G_{w=a}$ (sensing wall and agent as indiscriminate obstacles). Figure 3.11 reveals that our task environment required the ability to sense nearby animats and to differentiate between walls and animats in order to evolve reasonable EF values. Moreover, animats equipped with sensors on more sides achieved both, higher evolved fitness EF and higher reliability R across group sizes than the baseline setup $G_{0.50}$ and G_{random} (see Figure 3.11 and Figure 3.12(b)).

Overall, animats in the G_{3sides} condition consistently outperformed the animats in other groups except in two test conditions: *Blocked* and *Noisy Corners* (see Figure 3.6(a)). This shows that animats which are equipped with more sensors do have an advantage on average, but they may still perform worse than animats with fewer sensors under special circumstances (here: *Noisy Corners*). We assume that the sensory signals in these specific environments might have been too different from the information patterns the animats evolved in and were thus specialized for. Nevertheless, the additional sensors led to high reliability R across group sizes as well as relatively high task fitness for most modified wall-arrangements even though the animats evolved under a specific group size and a fixed wall configuration (see Figure 3.6(a,b)).

While $G_{w=a}$ animats had only one sensor which does not discriminate between the wall and other animats, $G_{no-agent}$ was missing the animat sensor completely. Still, $G_{no-agent}$ showed better task fitness than $G_{w=a}$ in test conditions with small group sizes and without a penalty. Considering the evolved behavior, $G_{w=a}$ animats (see Figure 3.12(c)) were not reactive to other animats, which suggests that they did not evolve the capacity to differentiate between the animats and the walls internally, e.g., through memory. While $G_{w=a}$ and $G_{no-agent}$ moved forward at similar rates, $G_{w=a}$ performed proportionally more turns than $G_{no-agent}$, which stood still more often.

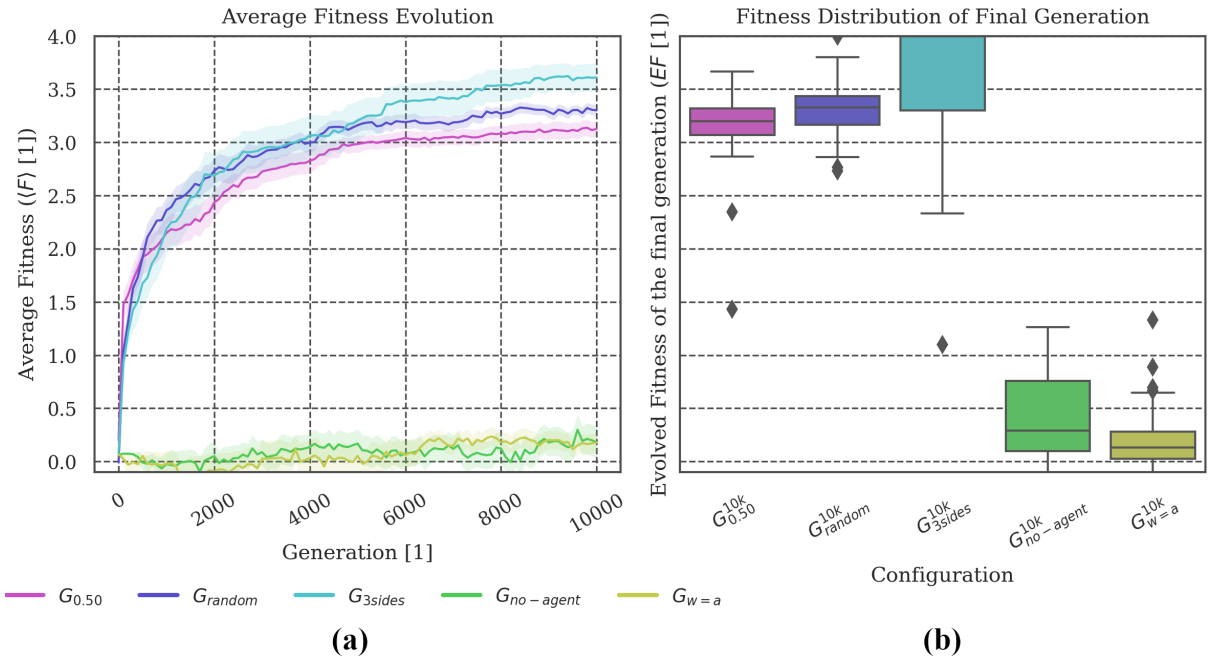


Figure 3.11: ***Fitness Evolution and distribution of the final evolved fitness.*** *The average evolved fitness showed that animats in evolutionary setups without specific sensors for other animats ($G_{no-agent}$ and $G_{w=a}$) achieved no reasonable fitness. By contrast, animats in G_{3sides} outperformed $G_{0.50}$, and G_{random}, but also had more outliers with lower fitness and performed worse than the baseline condition $G_{0.50}$ in early generations (up to $10,000$ generations).*

Analyzing the brain complexity showed that animats equipped with fewer, but also with more sensors than in the baseline setup $G_{0.50}$ evolved MBs with lower complexity (see Figure 3.13), albeit for different reasons. Based on the very low evolved fitness for $G_{w=a}$ and $G_{no-agent}$ (see Figure 3.11) we conclude that their MBs did not develop the necessary structure and mechanisms to solve the task, as reflected by their low brain complexity. By contrast, animats in G_{3sides} achieved high EF, $\langle TF \rangle$, and reliability R across group sizes, but did not evolve any integrated information ($\phi_{Max} = 0$) in most cases. This observation was in line with previous findings on the relation between sensory capacity and internal complexity (Albantakis et al. 2014) and suggested that high brain complexity in cognitive systems depends on a need for internal memory and computation, which may decrease if an animat is equipped with more sensors.

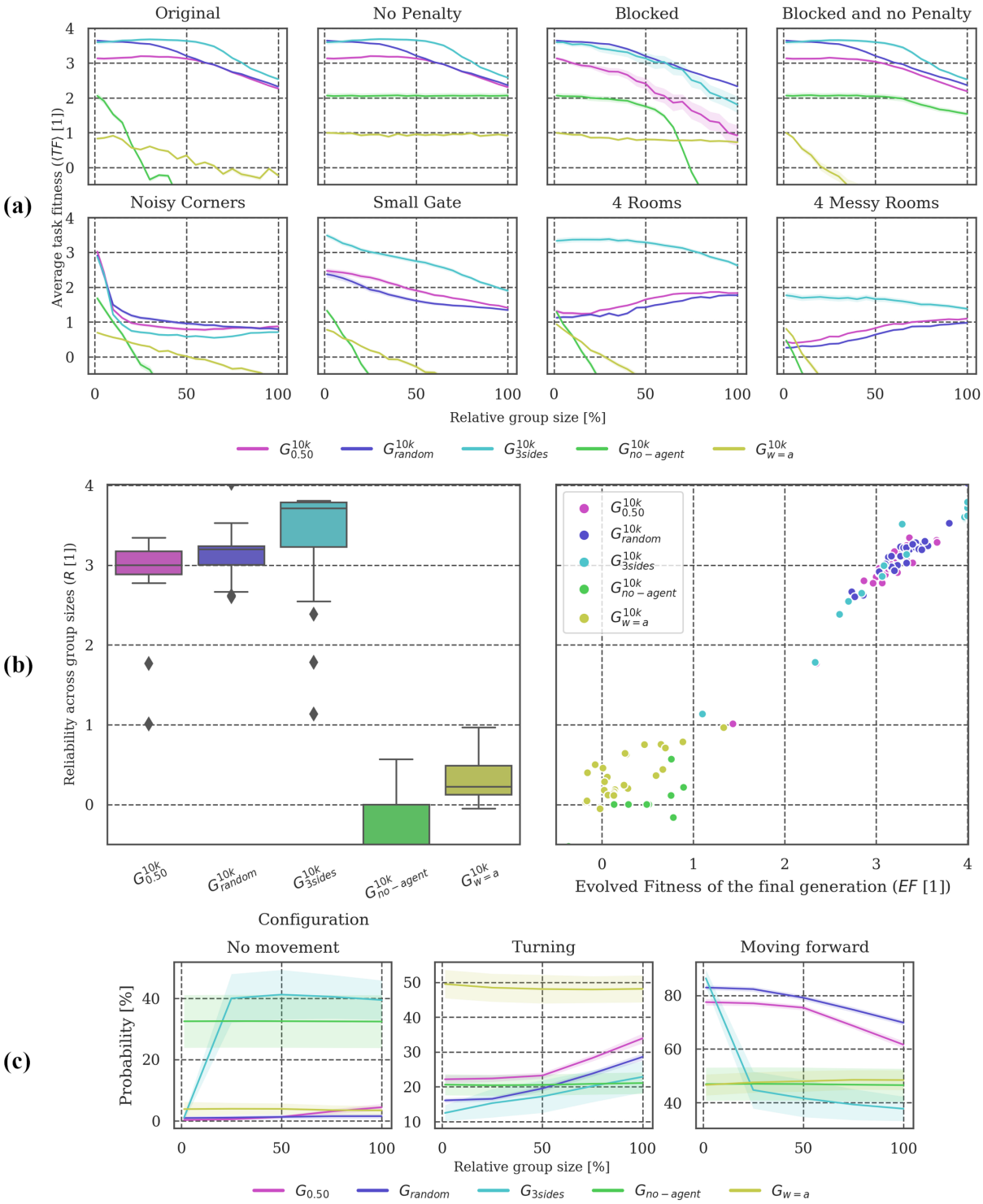


Figure 3.12: ***Post-evolutionary tests under modified conditions.*** *(a-b) The G_{3sides} condition had the highest $\langle TF \rangle$ in most test conditions, except in Blocked and Noisy Corners. In terms of R, sensing everything ($G_{w=a}$) with one sensor is still better than only sensing the walls ($G_{no-agent}$). (c) Setups with few sensors evolved no typical behavior (high variance of movement between the 30 different evolutions, shaded area). The G_{3sides} setup becomes more reactive as soon as the animat density starts to rise and thus evolved a different behavioral strategy than $G_{0.50}$ and G_{random}.*

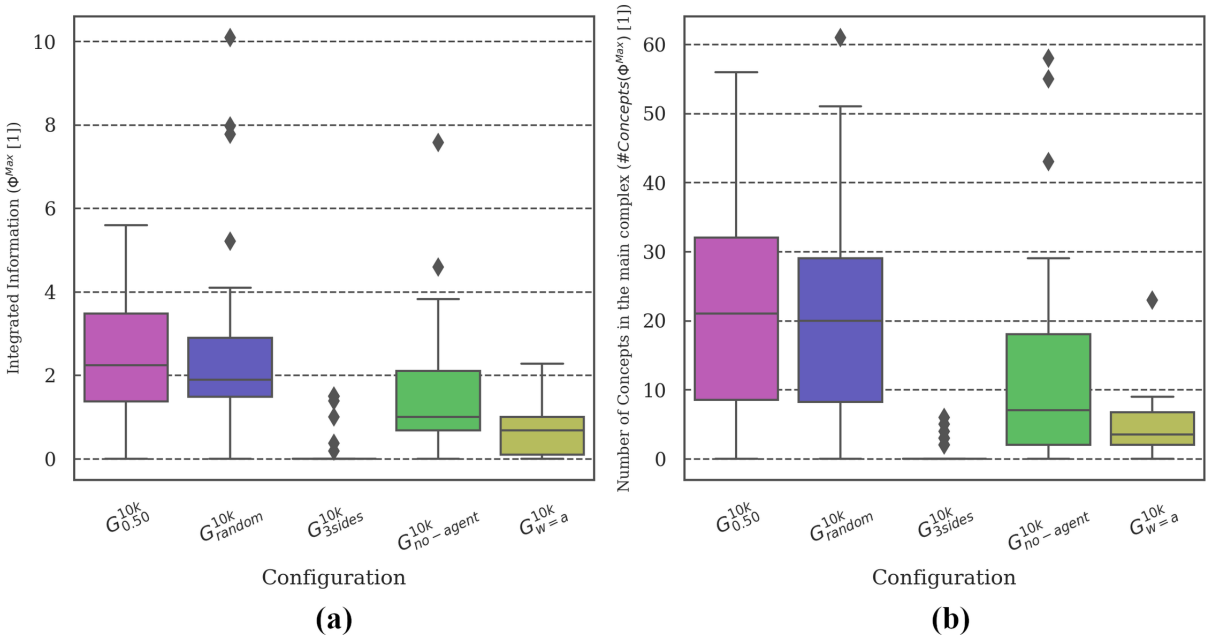


Figure 3.13: ***Distribution of brain complexity measures.*** *Animats in the G_{3sides} condition showed the lowest brain complexity of all setups despite having the highest evolved fitness and reliability. By contrast, animats with limited sensor information ($G_{no-agent}$ and $G_{w=a}$) had lower than baseline complexity values, but also low evolved fitness (EF).*

3.3 Discussion

The evolution of cooperative multi-agent systems might be the next frontier in the context of evolving artificial agents. To date, however, not much is known about conditions that give rise to cooperative behavior and the complex inter-dependencies between individual and group goals (Miikkulainen et al. 2012). For example, there might be many factors that influence whether the individuals either bow to the group or act by egoistic rules (Brown 1982). In this study, we used animats equipped with MBs, introduced by Edlund et al. (2011), to study how group performance and its reliability under modified conditions depends on the individual, interactions between individuals, as well as specific features of the MBs' evolution.

3.3.1 Prior work investigating group evolution

Earlier research that implemented groups of MBs concentrated on predator-prey environments and showed that animats can (co-)evolve swarm behavior (Olson 2015, Olson et al. 2012, 2013). The animat design in this work was generally based on a design in Marstaller et al. (2013), who evolved individual MBs with the goal of solving perceptual-categorization tasks. Another method of simulating swarm behavior is neuro-evolution, i.e., the evolution of Artificial Neural Networks (ANN) (Karpov et al. 2015, Stanley, Bryant & Miikkulainen 2005, Stanley, Cornelius, Miikkulainen, Silva & Gold 2005). As in Olson et al. (2012), these neuro-evolution experiments produced agents which evolve in a swarm to solve a predator-prey task.

Other researchers have investigated the effect of group size in the evolution of groups of simulated agents beyond predator-prey scenarios in a more general context. They find that the behavior of the group of agents and the individual agent is dependent on the group size (Hamann 2014, Garnier et al. 2007). In another study which changed the group size during evolution, the authors show that it can be easier for smaller groups than larger ones to organize themselves (Dorigo et al. 2004).

The effect of changing swarm sizes has also been investigated in the context of natural biological systems: Brown (1982) examined which factors are decisive for the individual to either join a swarm or behave egoistically. The study focused on experimenting with environmental qualities and swarm size. Brown defined optimal

swarm size as the best trade-off between the advantage of balancing costs between individuals in the swarm and the disadvantage of sharing the resources (energy/food) with the whole swarm. In an earlier study, Pacala et al. (1996) report that swarm size constrains information transfer and task allocation. They argue that the information exchange varies and the task allocation changes, depending on the swarm size of ant-colonies. Pacala et al. (1996) also argue that swarm behavior is the product of social interaction, individual interaction, and the interaction with the given environment. In a more recent work (Ishiwata et al. 2010), we found arguments that swarm behavior arises if there is sufficient density within the swarm.

3.3.2 Factors that impact evolved fitness and reliability

Generally, the ability to evolve high fitness in a given evolutionary setup depends on the interplay between external and internal factors as, e.g., the complexity of the environment and the animats' architecture, see also Albantakis et al. (2014). Exemplary for these factors, we manipulated the group size and the animats' sensorimotor and memory capacities across evolutionary setups. Further, we evaluated how these manipulations affected fitness evolution and post-evolutionary reliability.

Different group sizes. In the specific evolutionary setup investigated here, evolved fitness EF negatively correlated with group size, which is a result of the imposed penalty for collisions with other group members (see Figure 3.2 and 3.8, animats that evolved without the risk of penalty (G_{single} and $G_{no-penalty}$) achieved the highest $\langle EF \rangle$). On the other hand, animats evolved in fixed, intermediate group sizes (e.g., $G_{0.50}$ and $G_{0.25}$) are most reliable to changes in group size as measured by R, and, in fact, comparable to G_{random}, in which animats experienced random group sizes during evolution (see Figure 3.3(b)). The optimal group size for high R in our experiments is thus larger than the optimal group size for high EF, or individual fitness. This observation suggests, more generally, that unexpected changes in group size during evolution may sometimes lead to larger group sizes than expected based on what is best for an individual within the group.

Capacity for memory. Animats with less capacity for memory ($G_{smallbrain}$ and $G_{no-feedback}$) evolved to lower EF values than the baseline condition $G_{0.50}$ (see Figure 3.5). Further, the low memory setups were less reliable under changes in group size (low R). A higher memory capacity as in $G_{bigbrain}$ did not provide further advantages compared to $G_{0.50}$. Given the higher variance of $G_{bigbrain}$ in EF and R, we suspect that the larger search space made it more difficult for the evolutionary algorithm to converge to an optimal solution.

Sensorimotor capacity. Finally, more sensors (G_{3sides}) proved advantageous for both evolved fitness EF, reliability R across group sizes, and task fitness TF under almost all modified test conditions, including most modified wall arrangements (see Figure 3.12(b)). By contrast, training animats on multiple group sizes during evolution (G_{random}) led to high R, but did not translate to high task performance under modified wall arrangements (see Figure 3.3(b)). We speculate that the additional sensors allowed the animats to evolve more generalizable strategies in our two-dimensional spatial-navigation task, even though they evolved in a single static environment.

Note that we did not include a comparison condition in which animats evolved under various wall-arrangements, since it is not trivial to determine a statistically representative sample of all possible environments as part of the evolutionary simulation. For the same reason, we did not quantify average reliability across modified wall-arrangements, but provided task fitness measures for each tested wall-arrangement (see Figures 3.3/3.6/3.9/3.12(a)). In addition, Table 3.7 in Section 3.7 lists $\langle TF \rangle$ values for all evolutionary setups and test environments evaluated in this study.

Overall, our findings suggest that, in general, animats that were well-equipped for dealing with their original task environment (and thus achieved high evolved fitness) also performed better under modified conditions that were never encountered during evolution. Within most evolutionary setups, reliability R was correlated with evolved fitness (see Figures 3.3/3.6/3.9/3.12 (b), right Panel). The only exceptions were G_{single} and $G_{no-penalty}$, which did not adapt to the behavior of other group members at all. The high evolved fitness in G_{single} and $G_{no-penalty}$ could thus be interpreted as a form of narrow intelligence. By comparison, intermediate group sizes led to a somewhat more general form of intelligence.

Nevertheless, our findings also show that evolutionary setups that seem less adapted (lower evolved fitness) overall may still have advantages under some special modifications. For example, animats evolved in larger groups ($G_{1.0}$ and $G_{0.75}$) or with less memory capacity ($G_{smallbrain}$ and $G_{no-feedback}$) performed better than $G_{0.50}$ under most modified wall-arrangements (see Figure 3.3/3.6(a), bottom row; see Table 3.7). On the other hand, even G_{3sides} performed worse than the baseline ($G_{0.50}$) in one of the modified test environments (*Noisy Corners*).

3.3.3 Interactions between individuals in the group

In this study, we did not explicitly implement any form of direct communication between animats. Nevertheless, we found that it was necessary for animats to perceive their fellow group members and to distinguish them from static obstacles to achieve reasonable evolved fitness EF and reliability R (see Figures 3.11 and 3.12, where both $G_{no-agent}$ and $G_{w=a}$ overall show low values). Moreover, we observed that evolved interaction strategies provided advantages under certain modified conditions: Animats that evolved without a collision penalty ($G_{no-penalty}$) performed worse in some of the modified environments, even if tested without receiving a penalty (see Figure 3.9(a), *4 (Messy) Rooms*). While animats in $G_{no-penalty}$ were equipped with an agent sensor, they had no incentive to interact with or "pay attention" to their fellow agents. By contrast, the task fitness in the *4 (Messy) Rooms* conditions typically increased with group size for animats that evolved in groups and received either a collision penalty (e.g., $G_{0.25} - G_{1.00}$) and/or could not pass other agents ($G_{blocked}$ and $G_{blocked/no-penalty}$) (see Figures 3.3(a) and 3.9(a)). This indicates that they may have used other agents for orientation or guidance, a form of implicit cooperation. Indeed, animats evolved in large groups ($G_{0.75}$ and $G_{1.00}$) showed higher task fitness than $G_{0.50}$ in these particular modified test environments (see Figure 3.3(a), bottom; see Table 3.7).

As we know from previous studies, swarm behavior in nature can be the result of simple reactions to local neighbors (Garnier et al. 2007, Reid et al. 2015). For example, it could be a good strategy to stay close to a group member without hitting it. Such evolved behavior may then provide additional fitness advantages under some modified conditions (as in the *4 (Messy) Rooms* test condition here). The observed instances of cooperative behavior can thus be viewed as an emergent phenomenon of the evolutionary process.

3.3.4 Relation between brain complexity, evolved fitness, and reliability

Previous studies applying measures of integrated information to adaptive animats equipped with MBs (Albantakis et al. 2014, Edlund et al. 2011, Joshi et al. 2013) have observed that, on average, ϕ_{Max} and related measures for brain complexity increase over the course of evolution, which correlates with increasing evolved fitness EF (see Table 3.8 in Section 3.7). Moreover, as demonstrated in Albantakis et al. (2014), this increase depends on the complexity of the environment relative to the animats' sensor capacity: MBs that evolved in environments which require more memory and internal computation developed higher average ϕ_{Max} values and a higher number of concepts.

For the evolutionary setups with the baseline animat architecture as in $G_{0.50}$, we found the highest values of ϕ_{Max} and $\#Concepts(\phi_{Max})$ for medium group sizes $G_{0.50}$, $G_{blocked}$, and for G_{random}. These setups were also among the most reliable across group sizes. [6] By contrast, significantly lower ϕ_{Max} values were found for G_{single} and $G_{no-penalty}$, the two setups in which task fitness during evolution did not depend on interactions with other animats. As argued above, G_{single} and $G_{no-penalty}$ thus effectively evolved within a simpler task environment than $G_{0.50}$, $G_{blocked}$, and G_{random}, which explains their lower brain complexity ϕ_{Max}.

Compared to $G_{0.50}$, evolutionary setups with altered animat architectures showed consistently lower values of ϕ_{Max} and $\#Concepts(\phi_{Max})$. Limiting the animats' sensor capacity ($G_{no-agent}$ and $G_{w=a}$) or the number of available memory units ($G_{smallbrain}$ and $G_{no-feedback}$) interfered with their capacity for successful evolution in the spatial navigation task. Their lower evolved fitness was thus accompanied by less developed MBs with lower ϕ_{Max} and fewer concepts. Given more time to evolve (more generations), both their performance and their brain complexity might still increase. By contrast, more sensors allowed for better performance (EF, TF, and R) based

[6] See also Fischer et al. (2018) for similar results using a simplified measure of brain complexity.

on high amounts of external information, which effectively decreased the need for internal complexity (memory and computations) and thus may also lead to low ϕ_{Max}, as observed here for G_{3sides}.

In theory, high fitness in any given environment could be achieved without information integration ($\phi_{Max} = 0$) if no restrictions are imposed on the animats' architecture.[7] Moreover, information integration can be high even if there is no reasonable fitness, which partially explains the large variance in the brain complexity measures.[8] However, given a certain requirement for memory and context sensitivity, constraints in the number of sensors and memory elements may give rise to an empirical lower boundary on the amount of integrated information necessary to perform a given task (Albantakis et al. 2014, Edlund et al. 2011, Joshi et al. 2013, Sheneman & Hintze 2017).

In summary, for a given MB architecture, higher brain complexity seems to be related to better performance and reliability. However, future work should explore under which environmental conditions additional sensors, or more internal units, become more advantageous for the evolution of higher fitness (EF) and reliability (R).

3.3.5 Limitations

Our work modeled one particular, small-scale scenario of a multi-agent evolutionary setting. Future work should consider other types of environments which may strengthen the generality of our results. Moreover, further evolution or training scenarios for artificial organisms should be considered as well—here we do not use crossover in the genetic algorithm, for example, and all animats placed in the same environment are clones. In addition, Markov Brains are just one type of computational substrate and it would be interesting to see whether other types of substrates (e.g., Artificial Neural Networks) behave differently under modified test conditions (Hintze et al. 2019). Nevertheless, the results obtained in our simulation study could also be directly compared against certain types of biological models.[9]

While the measures that we employed to assess the complexity of the evolved MBs are theoretically motivated (Oizumi et al. 2014), they are also computationally very complex. This made it difficult to evaluate a larger sample size (number of evolution simulations) or to analyze the brain complexity of more generations (not only the final one). This is why alternative, approximate measures should be considered, too. For instance, the largest strongly connected component (and other graph metrics) can be used as a proxy for system integration and thus brain complexity (Fischer et al. 2018). Efficient approximations would also enable investigations into how brain complexity develops across generations as performed in (Albantakis et al. 2014) for slightly smaller MBs. Moreover, ϕ_{Max}, and the associated number of concepts $\#Concepts(\phi_{Max})$, are causal measures that assess the degree to which the mechanisms within a MB are differentiated and integrated. Future work should also consider and explore alternative informational or dynamical measures.[10] In this study, we concentrated on changes in task fitness and reliability under modified conditions, so the brain complexity analysis was not the subject of more in-depth investigation.

3.4 Conclusion

It is challenging to remain reliable in a dynamic and volatile world while also trying to succeed in a given task. Investigating the characteristics of this reliability, especially with regards to cooperative behavior, might also be useful to develop implications and strategies for improving the reliability of individuals within larger organizations. Despite complex dependencies between the individual, the group, and the environment, our computational approach offers a way to investigate reliability in group behavior. Here, we were particularly interested in the question of how cognitive and environmental constraints influence the reliability of simulated animats in a group. We were able to isolate essential influencing factors to better understand possible positive and negative effects of changing group size, environment design, and individual cognitive ability on reliability and task fitness under modified conditions. In particular, our study suggests that balancing the number of individuals in a group may lead

[7] E.g., by a system with a large feed-forward architecture (Oizumi et al. 2014).
[8] See, e.g., outliers for $G_{no-agent}$ in Figure 3.13.
[9] E.g., investigating the behavior of army ants under environmental modifications (Ishiwata et al. 2010, Reid et al. 2015).
[10] E.g., Beer & Williams (2015), Lizier et al. (2014), Zenil (2009).

to higher reliability under unforeseen changes in group size, even if the task itself would be simpler with fewer group members.

Moreover, a minimal number of sensors, the ability and incentive to distinguish static obstacles from other group members, and a minimal number of memory units were required to achieve high evolved fitness and reliability in our specific evolution simulations. If these minimal requirements were met, reliability R across group sizes was found to correlate with evolved fitness across the tested evolutionary setups. Limited sensor information forced the animats to evolve more complex brain structures, especially for intermediate group sizes, which also demonstrated the most reliable behavior across group sizes. Nevertheless, the highest task fitness across most modified conditions (varying group sizes as well as modified wall-arrangements) was observed for the evolutionary setup with additional sensors, which did not require high internal complexity. Finally, we presented data that support the evolution of implicit cooperation between animats. In all, this research asserts that task efficiency and effectiveness is not the only goal in dynamic environments; task reliability is also worth striving for.

3.5 Materials and methods

We used an evolutionary algorithm to generate simulated animats evolving in groups under various evolutionary setups (see Table 3.1), testing different animat architectures and evolutionary conditions to evolve animats having heterogeneous behavior, evolved fitness, and reliability. Afterwards, we conducted post-evolutionary tests to assess the reliability of the different evolutionary setups under modified conditions (see Table 3.2). This section explains the animat designs, the environment, the evolutionary simulations, and the experiment setup. We used Modular Agent Based Evolver (MABE) (Clifford Bohm, Nitash C. G. 2017) as a computational evolution framework with the same parameters as in previous work (Fischer et al. 2018) (see Table 3.6 in Section 3.7).

We chose MBs as a simplified model of an artificial brain, since the basic idea of an MB is to emulate the recurrent connectivity structure found in real neural networks in a simple manner, while being complex enough to represent a cognitive system (Marstaller et al. 2013). Furthermore, a recent study showed that MBs can be very compatible against variations of artificial neural networks and even showed higher performance in general (Hintze et al. 2017). Nevertheless, it would, in principle, also be possible to use a finite state machine (König et al. 2009), or artificial neural networks (Stanley, Cornelius, Miikkulainen, Silva & Gold 2005) to solve the kind of task investigated here.

Individual animats had to solve a two-dimensional spatial-navigation task in the presence of other animats (clones), thus forcing individuals to react to these other animats in order to reach a high fitness value. This task was a redesign by Fischer et al. (2018) of a task environment initially developed by König et al. (2009). An animat can usually differentiate between static (borders and walls) and dynamic objects (animats) in the environment through two distinct sensors. This design allowed for the evolution of social behavior[11] based on passive interactions between animats.

3.5.1 Animat architecture

The evolutionary algorithm evolves animats with MBs, which contain a set of discrete, binary computational units (”neurons”). Each unit has its own update rules receiving inputs from and sending their output to other units. In this study, the decision system (the connectivity between units and their update-rules) was implemented by Hidden Markov Gates (HMGs), which are encoded in an animat’s genome (string of integers $[0 - 255]$ with a minimum length of 2,000 elements and a maximum length of $20,000$ elements). The HMGs connect the nodes of the MB indirectly. Figure 3.14 visualizes a simple example, in which an HMG is connected to four units. The decision system inside an HMG can be diverse. In this research, we evolved discrete, deterministic lookup tables. The lookup tables translate the states of the connected input units at t to the new states of connected output units at

[11] We observed, e.g., ”waiting”, or ”following” behavior.

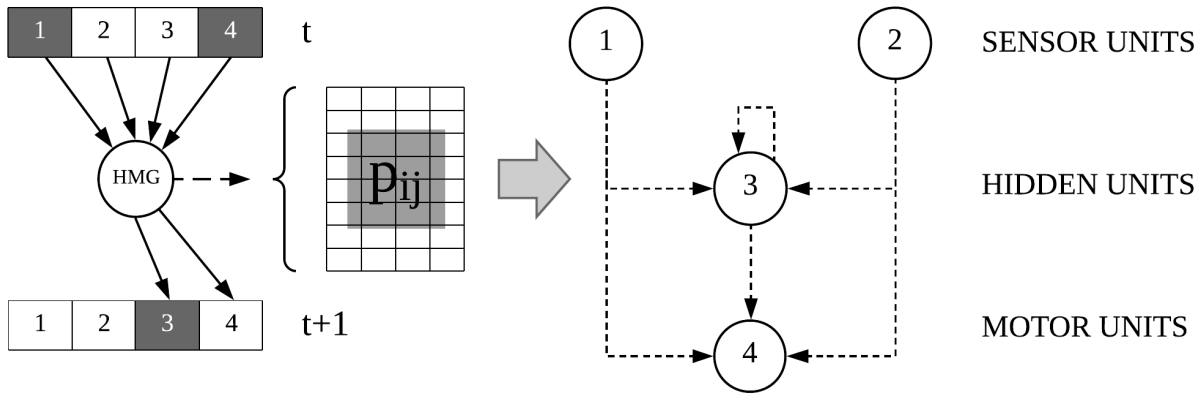


Figure 3.14: ***Example of a MB.*** *A MB (Edlund et al. 2011) has three components: (1) Units with a binary state ("*$1" - "4$*"), (2) HMGs and (3) the connections between the binary units and the HMGs. The connections between the units can be derived from the connections to the HMGs. HMGs contain the mechanism, e.g., a lookup table (here deterministic), to transform the brain state of units at* t *to the state at* $t + 1$.

$t + 1$. The motor or memory units can represent the output units of the HMG. The states of the sensor units are set by the input they receive from the environment.

The integers in an animat's genome encode the HMGs: the number of HMGs, their lookup tables, the connected input units, and the connected output units. The MBs evolve by mutating the genome in each new generation.[12] Each locus in the genome mutated with a certain probability (point mutations). In addition, larger sections could be deleted or added to the genome (Edlund et al. 2011, Hintze et al. 2017).[13] We did not use crossover or recombination (more than one parent per genome), since this would make it more difficult to trace an animat's line of descent without obvious computational advantages in the simple evolutionary setting investigated here. In principle, other optimization algorithms could be employed to develop well-performing MBs. The evolutionary algorithm used here has the advantage that both the node connectivity and the nodes' update rules can be encoded in the genome and jointly adapted through mutation and fitness selection.

All units in the animat's MB have binary states, either 1 or 0. A sensor turns 1 if an obstacle is detected and a motor switches to 1 if it is active. Two motors provide the ability to turn 90 degrees left or right, and to move forward (if both motors are in state 1). Since the units within a MB can be interconnected in a recurrent manner, they have the potential to create internal memory. We evolved animats with five different animat designs displayed in Figure 3.15. The baseline cognitive architecture was introduced already in Fischer et al. (2018).[14] Here, further deviations were designed to investigate the influence of an animat's sensorimotor and memory capacities on the resulting evolved fitness and the animats' task fitness and reliability under modified post-evolutionary test conditions. The sensors had a detection range of one unit. Typically, the motor units could also feedback to the memory and motor units, thus acting as additional memory capacity, since knowledge about previous motor states is directly available for computing the next state. One animat design was included that lacked the possibility for motor feedback ($G_{no-feedback}$).

3.5.2 Design of the 2D environment

All experiments simulated a two-dimensional environment. The world has 32×32 units (see Figure 3.16). All animats started on one of 72 predefined, uniformly distributed, starting positions. The selection for the starting position, as well as an animat's initial orientation, was random at every new generation. The original environment (see Figure 3.16(a)) had two rooms, which are connected by a gate. The animats' goal was to travel between the

[12] See Olson et al. (2013) and Hintze et al. (2019)
[13] Again, all parameters are listed in Table 3.6 within Section 3.7.
[14] The animats are equipped with one front wall sensor, one front agent sensor, four memory units, and two motors.

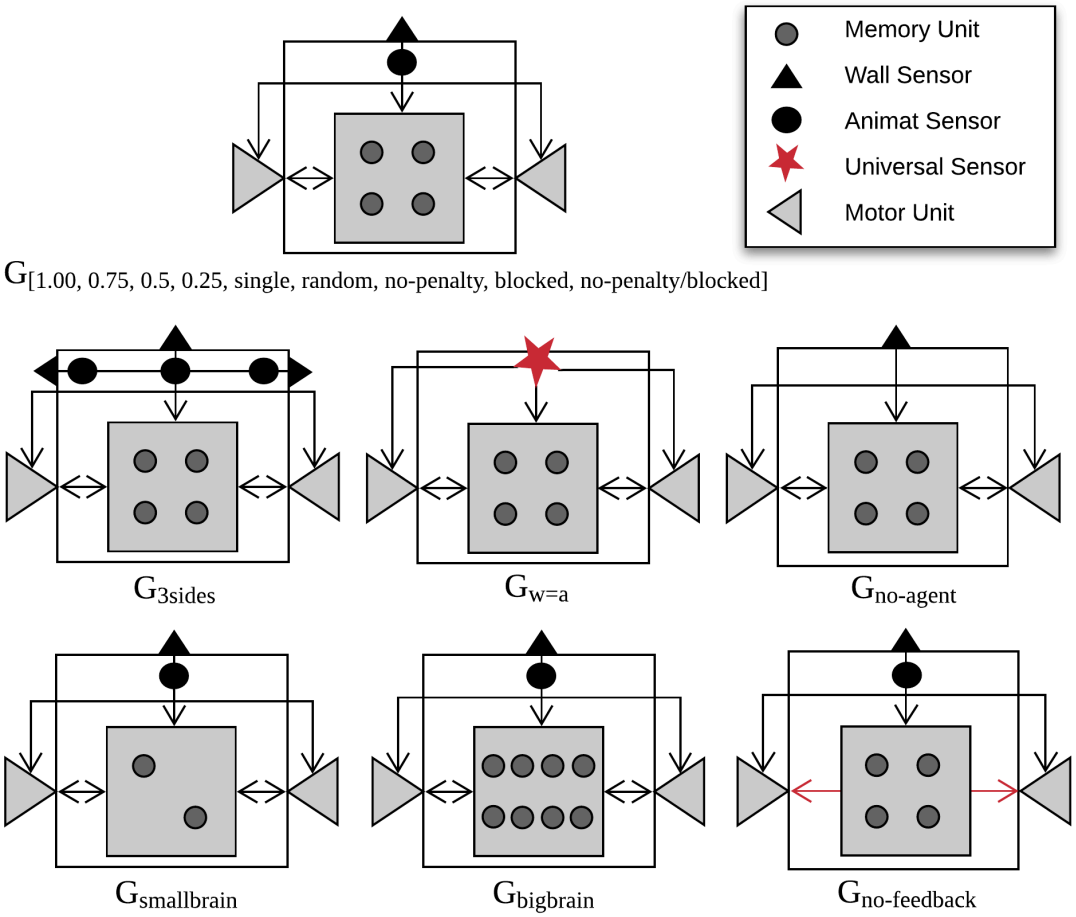


Figure 3.15: ***Schematic architecture of the five different animat architectures.*** *The top row shows the original animat architecture as defined in (Fischer et al. 2018). The animats have two motor units (grey triangles), four memory units (dark grey circles) and one to six sensor units (black/red shapes). The middle row shows animats with a changed sensor architecture, from the left: The architecture with sensors on three sides, the architecture with a single sensor unit, detecting wall and animat indiscriminately, and the architecture without an animat sensor. The bottom row shows animats with changed memory architecture, from the left: The architecture with only two memory units, the architecture with eight memory units and the architecture without feedback motors (motors cannot be part of the memory network). Note that the architectures depict the maximal number of units available. Whether any given unit is actually used depends on the evolved connectivity and logic function. Animats are initialized in the first generation without connections between units.*

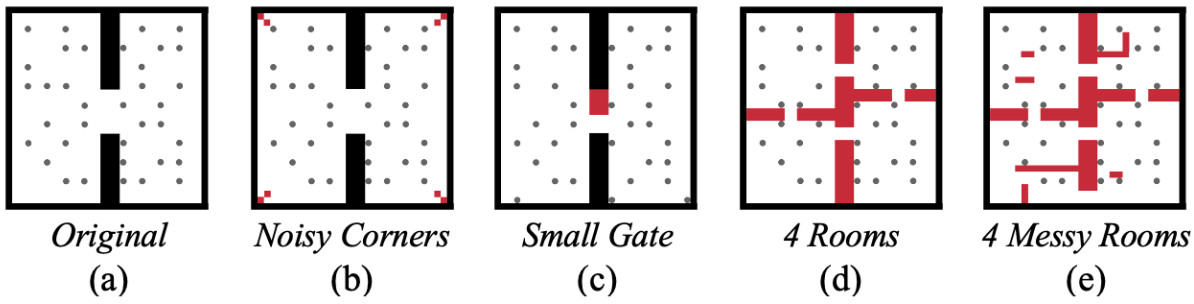


Figure 3.16: ***Environmental design.*** *(a) The two-dimensional environment is based on a discrete grid architecture and contains two rooms. Animats draw a random starting position. Their orientation can be up, down, left, and right and is also randomly selected at initiation. (b-e) Four additional environments were used to test the task fitness of the animats under modified conditions. Red blocks mark the changes/additions in the room and represent walls. In (d), all four gates count as possible rewards. In (e), only gates on the vertical mid-line provide rewards.*

two rooms in order to achieve a high fitness value. This design was adapted from the work of König et al. (2009). All evolutionary setups evolved in the original environment. As an additional test dimension for evaluating task fitness under modified conditions, we tested all evolved MBs (the final generation) in four modified environment designs (see Figure 3.16(b-e)). Generally, animats were allowed to inhabit the same location in the environment (albeit under penalty, see below), except in $G_{blocked}$ and $G_{blocked/no-penalty}$.

3.5.3 Experiment design

We selected $G_{0.50}$ to be the baseline setup for evolution, to which we compared all other evolutionary setups. This was because $G_{0.50}$ showed the highest reliability R across group sizes. In sum, we came up with 15 different setups for the evolution of the animats (see Table 3.1). Using the MABE framework, we simulated each evolutionary setup 30 times. In each of these 30 evolutions, the evolutionary algorithm had $10,000$ generations to converge on the final solution. A population of 100 genomes was mutated and evaluated in each generation. Each of these evaluations was repeated 30 times (30 "test runs") with random starting positions, orientation, and selection order for simulating the animats movement serially. Random seeds were chosen using a Mersenne-Twister (`mt19937`) random number generator.[15] After a genome was tested 30 times, it received a fitness score, which was computed based on the mean across the task performance of 30 single animats, with one being picked randomly from each of the 30 random test runs. In addition, in setup G_{random} the group size varied for each of the 30 tests. The specific group size was drawn randomly from a static vector.[16] This vector simulates a uniform distribution between 1 and 72.

3.5.4 The simulated life

The fitness function F that determines the probability of a genome being reproduced depends on two factors. First, animats A have to travel as often as possible through the gate (change the room, see Figure 3.16). Second, the animats need to avoid colliding with each other. Fischer et al. (2018) already included the formal definitions of the fitness function as a weighted sum of the penalty for collision and the reward for crossing the gate (see Table 3.5 for the mathematical notation of equations 3.4 and 3.5):

$$f(a) = \sum_{t=0}^{T-1} \begin{cases} 1 & g(a,t,t+1) = 1 \text{ and } g(a,t-100,t) = 0 \\ 0 & \textit{otherwise} \end{cases} - \sum_{t=0}^{T} \begin{cases} 0.075 & c(x(a),y(a),t) > 1 \\ 0 & \textit{otherwise} \end{cases} \tag{3.4}$$

[15] See Section 3.7 for a more detailed explanation of the parameter sampling.
[16] $[1, 4, 7, 11, 14, 18, 22, 25, 29, 32, 36, 40, 43, 47, 50, 54, 58, 61, 65, 68, 72]$

Table 3.5: *Mathematical notation as used in the fitness function F(A) and f(a).*

$a \in A$	A single animat a in the set of all animats A in a trial.
$f(a)$	The fitness of a single animat a.
$F(A)$	The average fitness of all animats in A as clones of a single genome.
$rand(A)$	Picks a random animat a from the group A.
$g(a, t_a, t_b)$	Returns the number of gate-crossings between time ta and time t_b for a single animat a.
$t \in T$	A single time step t, where $t \in T$ and $T = [1, 2, ..., 499, 500]$.
$c(x, y, t)$	Returns the number of animats at a specific position (x, y) at time t.

$$F(A) = \frac{\sum_{i=1}^{30} f(\text{rand}(A))}{30} \tag{3.5}$$

The amount of reward ($+1.0$ points) is higher than the amount subtracted in the case of a penalty (-0.075 points). These numbers need to be chosen carefully. If the penalty is too low or the reward is too high, animats will keep moving from one room to the other through the gate (herding effect) and ignore the penalty. On the other hand, given a high penalty and low reward, animats will evolve hardly any movement. To further reduce the herding effect around the gate, there is a refractory period of 100 timesteps after receiving a reward before the same animat can receive another reward. Since each trial has a duration T of 500 timesteps, any one animat can receive a total fitness score of at most 4 points (Fischer et al. 2018).

To investigate the coordination and cooperation of animats in groups, we let animats co-exist in the same environment.[17] Currently, we have not implemented co-evolution of animats with different genomes and have only evaluated a genome by generating animats as identical clones (with the same MBs). There was no active knowledge exchange (”communication”) between animats in this study. Animats had to develop the ability to distinguish which kind of sensory input to use for decision making. As specified above, sensors can only sense one position in front of – or on the side of (G_{3sides}) – the animat and differentiate between static objects (walls) and dynamic objects (fellow animats), except for $G_{w=a}$.

Compared to the baseline setup, we included further evolutionary setups in which animats did not receive the collision penalty and/or were not able to overlap ($G_{no-penalty}$, $G_{blocked}$, $G_{blocked/no-penalty}$). Those changes in the fitness function represented environmental rules which influenced the task difficulty. As a result, we were able to test the role that the imposed interaction conditions between animats played in order to achieve high task fitness under modified conditions.

3.5.5 Post-evolutionary evaluation

Modified conditions. Post-evolutionary task fitness tests were designed as follows: First, we selected the 30 genomes of generation $10,000$ $(10k)$ for each of the 15 evolutionary setups (see Table 3.1). Second, each genome was tested across 21 conditions varying in group size in the *Original* test condition. To this end, we created groups of animat clones of the respective test group size for each of the 30×15 genomes. Test group sizes were uniformly distributed between *one*[18] and 72. The interval of the relative group sizes is [1, 4, 7, 11, 14, 18, 22, 25, 29, 32, 36, 40, 43, 47, 50, 54, 58, 61, 65, 68, 72].

In addition to varying group sizes in the baseline task design (*Original*), we created four modified test environments, as shown in Figure 3.16 (*Noisy Corners, Small Gate, 4 Rooms, 4 Messy Rooms*). Moreover, we included three additional test conditions in which we varied the interaction conditions of the animats (*No Penalty, Blocked, Blocked and no penalty*). Finally, we tested each of the $30 \times 15 \times 21$ different configurations in each of the eight test environments.

For the statistical analysis and the main reliability evaluations, we defined a quantitative reliability measure R across group sizes in the Original environment design (see equation 3.3 above). The modified test environments

[17] In contrast to previous studies in this scope (Marstaller et al. 2013, Oizumi et al. 2014, Edlund et al. 2011).

[18] A single animat is not a group, but we treat it as one in order to simplify notation.

represented four independent samples of possible environmental modifications. For this reason, they were evaluated on their own in terms of the achieved TF. The results of the remaining three test conditions with varying interaction properties mainly served to highlight differences between the evolutionary setups, rather than testing reliability per se.

Brain complexity. To evaluate the complexity of the evolved MBs, we employed two complimentary measures provided by IIT (Oizumi et al. 2014, Tononi & Koch 2015, Tononi 2015): ϕ_{Max} and the associated number concepts $\#Concepts(\phi_{Max})$. The core of IIT's measures is an information theoretic, and probabilistic graph analysis (Oizumi et al. 2014) based on the state-to-state transition probabilities of the units, i.e., their update functions. Please refer to Oizumi et al. (2014) and Albantakis et al. (2014) for details on the evaluation. Very briefly, to evaluate the integrated information ϕ ("big phi") for a particular set of computational units S in state $S = s$, the first step is to assess which subsets $Y \subseteq S$ specify positive integrated information $\varphi > 0$ ("small phi") within the system (the set's "concepts"). φ captures how much a set of elements Y within the system in its state y constrains the prior and next states of other system subsets $V_{(t\pm1)} \subseteq S$. In simplified terms:

$$\varphi(Y = y_t) = \min_{t\pm1} \left(\min_{\Psi} \left(D \left(\frac{\hat{p}(V_{t\pm1}|y_t)}{\Psi(\hat{p}(V_{t\pm1}|y_t))} \right) \right) \right) \tag{3.6}$$

where Ψ partitions $\hat{p}(V_{t\pm1}|y_t)$ into the product distribution $\hat{p}(V_{1,t\pm1}|y_{1,t}) \times \hat{p}(V_{2,t\pm1}|y_{2,t})$, and D is a distance measure between two probability distributions. The ^ ("hat"-symbol) above the probability function p indicates that probabilities are interventional (obtained from system perturbations) rather than observational (Oizumi et al. 2014, Ay & Polani 2008). $V_{t\pm1}$ are chosen such that $\varphi(y_t)$ is maximal. Second, ϕ is measured as the minimal difference that any system partition Ψ_S makes to the overall information specified by all subsets Y with $\varphi(y_t) > 0$. Again, in simplified terms:

$$\phi = \min_{\Psi_S}(D(\{\varphi(y_t)\}_{Y\subseteq S}; \Psi_S(\{\varphi(y_t)\}_{Y\subseteq S}))) \tag{3.7}$$

For a given MB, we search across all sets of computational units S for the one with $\phi_{Max} = \max_S \phi$. ϕ_{Max} represents the highest possible integrated information the MB can achieve across all its subsets, which we used as an indicator for brain complexity (Oizumi et al. 2014).

All calculations were conducted using the IIT Python package `pyphi` (Mayner et al. 2018), which we used in our work to calculate ϕ_{Max} and the corresponding number of concepts. Since the employed measures are state-dependent, we evaluated ϕ_{Max} and the number of concepts for every state a MB experienced during a lifetime (one trial) and selected the maximum value over all states as in (Albantakis et al. 2014). Figure 3.17 in Section 3.7 shows by way of example that it is essential for high ϕ_{Max} in a system that many elements are integrated, meaning also maintaining functional feedback loops within the system. In this study, we only considered the brain complexity of the final generation ($10k$) due to the computational complexity of calculations using `pyphi`.

Statistics. The evolved fitness values EF, the reliability R, and the IIT brain complexity measures were statistically evaluated across all evolutionary setups using a Kruskal-Wallis test, which showed a significant difference of the observed statistics between all groups taken together. Further, we used the Mann-Whitney-U test to evaluate the difference between pairs of evolutionary setups. Section 3.7 lists all statistical tests that are a subject of discussion in the results and discussion section.

3.6 References

Albantakis, L., Hintze, A., Koch, C., Adami, C. & Tononi, G. (2014), 'Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity', *PLoS Computational Biology* **10**(12), e1003966.
URL: *https://dx.plos.org/10.1371/journal.pcbi.1003966*

Ay, N. & Polani, D. (2008), 'Information Flows in Causal Networks', *Advances in Complex Systems* **11**(01), 17–41.

Beer, R. D. & Williams, P. L. (2015), 'Information Processing and Dynamics in Minimally Cognitive Agents', *Cognitive Science* **39**(1), 1–38.
URL: *http://doi.wiley.com/10.1111/cogs.12142*

Brown, J. L. (1982), 'Optimal group size in territorial animals', *Journal of Theoretical Biology* **95**(4), 793–810.

Clifford Bohm, Nitash C. G., A. H. (2017), MABE (Modular Agent Based Evolver): A framework for digital evolution research, *in* 'Proceedings of the European Conference on Artificial Life', MIT Press, pp. 76–83.

Dorigo, M., Trianni, V., Şahin, E., Groß, R., Labella, T. H., Baldassarre, G., Nolfi, S., Deneubourg, J.-L., Mondada, F., Floreano, D. & Gambardella, L. M. (2004), 'Evolving Self-Organizing Behaviors for a Swarm-Bot', *Autonomous Robots* **17**(2/3), 223–245.

Edlund, J. A., Chaumont, N., Hintze, A., Koch, C., Tononi, G. & Adami, C. (2011), 'Integrated Information Increases with Fitness in the Evolution of Animats', *PLoS Computational Biology* **7**(10), e1002236.

Engel, D. & Malone, T. W. (2018), 'Integrated information as a metric for group interaction', *PLOS ONE* **13**(10).

Fischer, D., Mostaghim, S. & Albantakis, L. (2018), 'How swarm size during evolution impacts the behavior, generalizability, and brain complexity of animats performing a spatial navigation task', *GECCO 2018* .
URL: *http://dx.doi.org/10.1145/3205455.3205646*

Fleck, L. (1983), *Genesis and Development of a Scientific Fact*, Vol. 11.

Gardner, H. (1987), 'The theory of multiple intelligences', *Annals of Dyslexia* **37**(1), 19–35.

Garnier, S., Gautrais, J. & Theraulaz, G. (2007), 'The biological principles of swarm intelligence', *Swarm Intelligence* **1**(1), 3–31.

Hamann, H. (2014), 'Evolution of Collective Behaviors by Minimizing Surprise', *14th Int. Conf. on the Synthesis and Simulation of Living Systems (ALIFE 2014)* pp. 344–351.

Hintze, A., Edlund, J. A., Olson, R. S., Knoester, D. B., Schossau, J., Albantakis, L., Tehrani-Saleh, A., Kvam, P., Sheneman, L., Goldsby, H., Bohm, C. & Adami, C. (2017), 'Markov Brains: A Technical Introduction'.
URL: *http://arxiv.org/abs/1709.05601*

Hintze, A., Kirkpatrick, D. & Adami, C. (2018), 'The structure of evolved representations across different substrates for artificial intelligence'.
URL: *http://arxiv.org/abs/1804.01660*

Hintze, A., Schossau, J. & Bohm, C. (2019), The Evolutionary Buffet Method, *in* 'Genetic Programming Theory and Practice XVI', pp. 17–36.
URL: *http://link.springer.com/10.1007/978-3-030-04735-1_2*

Ishiwata, H., Noman, N. & Iba, H. (2010), Emergence of Cooperation in a Bio-inspired Multi-agent System, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 6464 LNAI, pp. 364–374.

Joshi, N. J., Tononi, G. & Koch, C. (2013), 'The Minimal Complexity of Adapting Agents Increases with Fitness', *PLoS Computational Biology* **9**(7).

Karpov, I. V., Johnson, L. M. & Miikkulainen, R. (2015), Evaluating team behaviors constructed with human-guided machine learning, *in* '2015 IEEE Conference on Computational Intelligence and Games (CIG)', IEEE, pp. 292–298.

König, L., Mostaghim, S. & Schmeck, H. (2009), 'Decentralized evolution of robotic behavior using finite state machines', *International Journal of Intelligent Computing and Cybernetics* **2**(4), 695–723.

List, C. & Pettit, P. (2006), 'Group Agency and Supervenience', *The Southern Journal of Philosophy* **44**(May 2005), 1–22.

Lizier, J. T., Prokopenko, M. & Zomaya, A. Y. (2014), A Framework for the Local Information Dynamics of Distributed Computation in Complex Systems, pp. 115–158.
URL: *http://dx.doi.org/10.1007/978-3-642-53734-9_5*

Marshall, W., Albantakis, L. & Tononi, G. (2018), 'Black-boxing and cause-effect power', *PLOS Computational Biology* **14**(4), e1006114.
URL: *http://dx.plos.org/10.1371/journal.pcbi.1006114*

Marshall, W., Kim, H., Walker, S. I., Tononi, G. & Albantakis, L. (2017), 'How causal analysis can reveal autonomy in models of biological systems', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**(2109), 20160358.

Marstaller, L., Hintze, A. & Adami, C. (2012), 'The evolution of representation in simple cognitive networks', pp. 1–36.
URL: *http://dx.doi.org/10.1162/NECO_a_00475*

Marstaller, L., Hintze, A. & Adami, C. (2013), 'The Evolution of Representation in Simple Cognitive Networks', *Neural Computation* **25**(8), 2079–2107.

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R. & Tononi, G. (2018), 'PyPhi: A toolbox for integrated information theory', *PLOS Computational Biology* **14**(7), e1006343.

Miikkulainen, R., Feasley, E., Johnson, L., Karpov, I., Rajagopalan, P., Rawal, A. & Tansey, W. (2012), Multiagent Learning through Neuroevolution, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 7311 LNCS, pp. 24–46.

Nonaka, I. (2000), 'A firm as a knowledge-creating entity: a new perspective on the theory of the firm', *Industrial and Corporate Change* **9**(1), 1–20.

Oizumi, M., Albantakis, L. & Tononi, G. (2014), 'From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* **10**(5), 1–25.

Oliver, N., Senturk, M., Calvard, T. S., Potocnik, K. & Tomasella, M. (2017), 'Collective Mindfulness, Resilience and Team Performance', *Academy of Management Proceedings* **2017**(1).

Olson, R. S. (2015), Elucidating the Evolutionary Origins of Collective Animal Behavior, PhD thesis.

Olson, R. S., Hintze, A., Dyer, F. C., Knoester, D. B. & Adami, C. (2012), 'Predator confusion is sufficient to evolve swarming behavior', *Journal of the Royal Society, Interface / the Royal Society* **10**, 20130305.

Olson, R. S., Knoester, D. B. & Adami, C. (2013), 'Critical interplay between density-dependent predation and evolution of the selfish herd', *Proceeding of the fifteenth annual conference on Genetic and evolutionary computation conference - GECCO '13* p. 247.
URL: *https://doi.org/10.1145/2463372.2463394*

Pacala, S. W., Gordon, D. M. & Godfray, H. C. J. (1996), 'Effects of social group size on information transfer and task allocation', *Evolutionary Ecology* **10**(2), 127–165.
URL: *http://link.springer.com/10.1007/BF01241782*

Pinter-Wollman, N., Penn, A., Theraulaz, G. & Fiore, S. M. (2018), 'Interdisciplinary approaches for uncovering the impacts of architecture on collective behaviour', *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**(1753).

Reid, C. R., Lutz, M. J., Powell, S., Kao, A. B., Couzin, I. D. & Garnier, S. (2015), 'Army ants dynamically adjust living bridges in response to a cost-benefit trade-off.', *Proceedings of the National Academy of Sciences of the United States of America* **112**(49), 15113–8.

Sheneman, L. & Hintze, A. (2017), 'Evolving autonomous learning in cognitive networks', *Scientific Reports* (July), 1–11.
URL: *http://dx.doi.org/10.1038/s41598-017-16548-2*

Spearman, C. (1904), '"General Intelligence," Objectively Determined and Measured', *The American Journal of Psychology* **15**(2), 201–292.

Stanley, K. O., Bryant, B. D. & Miikkulainen, R. (2005), 'Real-time neuroevolution in the NERO video game', *IEEE Transactions on Evolutionary Computation* **9**(6), 653–668.

Stanley, K. O., Cornelius, R., Miikkulainen, R., Silva, T. D. & Gold, A. (2005), 'Real-time Learning in the NERO Video Game', *Proceedings of the First Artificial Intelligence and Interactive Digital Entertainment Conference* **2003**, 2003–2004.

Tononi, G. (2015), 'Integrated information theory', *Scholarpedia* **10**(1), 4164. revision #150725.

Tononi, G. & Koch, C. (2015), 'Consciousness: here, there and everywhere?', *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**(1668), 20140167–20140167.

Tsoukas, H. (1996), 'The firm as a distributed knowledge system: A constructionist approach', *Strategic Management Journal* **17**(S2), 11–25.

Walsh, J. & Ungson, G. R. (1991), 'Organizational Memory', *Academy of Management Review* **16**(1), 57–91.

Weick, K. E. & Roberts, K. H. (1993), 'Collective Mind in Organizations: Heedful Interrelating on Flight Decks', *Administrative Science Quarterly* **38**(3), 357.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2008), 'Organizing for High Reliability: Process of Collective Mindfulness', *Crisis Management* **3**, 31–66.

Zenil, H. (2009), 'Compression-based investigation of the dynamical properties of cellular automata and other systems', *Arxiv preprint arXiv:0910.4042* pp. 1–25.
URL: *http://arxiv.org/abs/0910.4042*

3.7 Supporting information

3.7.1 Brain wiring diagram

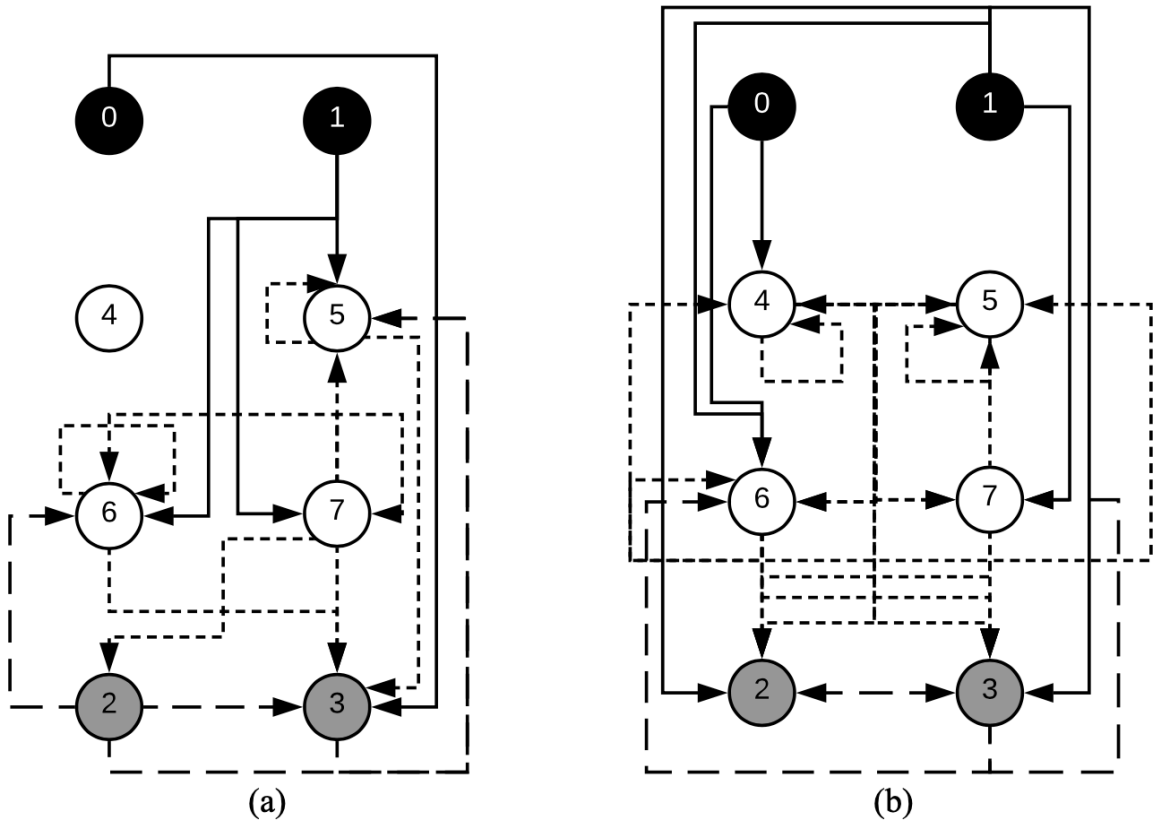


Figure 3.17: ***Brain wiring diagram.*** *(a). Best animat in evolution #4 under condition G_{random} with an evolved fitness $EF = 3.1$ and $\phi_{Max} = 0$. The network structure shows only few feedback loops, which cannot produce integrated information. (b) Best animat in evolution #1 under condition G_{random} with an evolved fitness $EF = 2.9$ and $\phi_{Max} = 7.77$. The network structure shows much more connections, which integrated the network states and makes them interdependent.*

3.7.2 Parameters for the genetic algorithm

Table 3.6: *Parameters used to configure the Genetic Algorithm with in the MABE framework.*

Category	Setting	Value
Genome setup	Type	Circular
	Alphabet Size	256
	Sites Type	char
	Initial Size	5
	Mutation Point Rate	0.005
	Mutation Copy/Delete Rate	0.00002
	Minimal Mutation Copy/Delete Size	128
	Maximum Mutation Copy/Delete Size	512
	Minimal Size	2
	Maximal Size	20
Markov brain setup	Type of Gates	Deterministic
	Range of Inputs/Outputs per Gate	1 to 4
Optimizer setup	Type of Optimizer	Tournament
	Tournament Size	5
	Population Size	100
	Elitism	No
	Number of Parents	1 (no crossover)

3.7.3 Parameter sampling

Description of the random seeds and random number generator used in this study:

We used a Mersenne-Twister (`mt19937`) random number generator throughout the simulation. Each experiment condition G_i was evolved 30 times on 30 distinct random seeds (corresponding to the $N = 30$ different evolution simulations per evolutionary setup G_i). The set of 30 distinct random seeds was manually chosen to make the experiments reproducible. The random number generator was used to draw a set of 30 unique samples for the starting positions as each genome was evaluated 30 times per generation. In addition, the random number generator was used to draw 30×500 samples per seed for the selection order for each simulation trial, which determines the order in which the individual animats updated their position and orientation in each of the 500 time steps per seed. In other words, the animat's perception and reaction was processed serially under a random sequence. Note that the update sequence per seed remained the same across all $10,000$ generations per simulation trial, while the distribution of the starting orientation (up, down, left, right) of the animats was drawn continuously, differing for each evaluation. All numbers were drawn from a uniform distribution. To reduce biases produced by the random number generator, we performed the post-evolutionary tests on different machines and random seeds. This means that initial conditions differed between the evolution simulations and the post-evolutionary tests.

3.7.4 Statistical tests

Table 3.7: *Mean ⟨...⟩ and SEM values of the evolved fitness EF, reliability across group sizes in the original setup R, and brain complexity ϕ_{Max} of the final evolved animats grouped by conditions. Roman numbers indicated the rank of the corresponding mean through all conditions. The results of brain complexity calculations for $G_{bigbrain}$ are not available (NA) due to the computational complexity of the calculations.*

G	EF	SEM	R	SEM	ϕ^{Max}	SEM
0.50	(VIII) 3.1267	0.0727	(III) 2.9315	0.0845	(III) 2.4921	0.2673
random	(VI) 3.3051	0.0492	(II) 3.0257	0.0491	(II) 2.6472	0.4319
1.00	24.271	0.0822	(VIII) 2.3580	0.1415	(VII) 1.8157	0.2595
0.75	26.732	0.0891	(VII) 2.6239	0.1039	(VI) 1.9209	0.3467
0.25	(V) 3.3099	0.1188	(VI) 2.7311	0.0945	(V) 2.1980	0.3278
single	(II) 3.7356	0.0588	-22.667	0.6666	(X) 1.2476	0.1697
bigbrain	(VII) 3.1847	0.1017	(IV) 2.7725	0.1047	NA	NA
smallbrain	(X) 2.7296	0.0617	(IX) 2.3343	0.0809	(IX) 1.1600	0.1906
no-feedback	19.696	0.1131	(X) 1.7839	0.1165	0.3389	0.0824
no-agent	0.1788	0.1474	0.3168	0.1865	(VIII) 1.5529	0.3064
3sides	(III) 3.6092	0.1281	(I) 3.3808	0.1227	0.2483	0.0960
w=a	0.1793	0.0881	0.2632	0.0868	0.6721	0.1025
no-penalty	(I) 3.7867	0.0447	-28.575	0.2115	0.9356	0.1348
blocked/no-penalty	(IV) 3.3367	0.0532	0.5958	0.2988	(IV) 2.2055	0.2271
blocked	(IX) 3.0994	0.0924	(V) 2.7442	0.0806	(I) 3.0634	0.5598

Table 3.13: *Task fitness values of all conditions G_i in the five evaluated environments.*

TF	Map m =				
	Original	**Noisy Corners**	**Small Gate**	**4 Rooms**	**4 Messy Rooms**
0.25	2.71	0.98	1.50	1.54	0.80
random	3.02	0.94	1.60	1.50	0.70
0.50	2.89	0.84	1.83	1.66	0.87
0.75	2.66	1.08	1.86	1.94	1.06
1.00	2.48	0.96	1.78	1.82	1.13
single	-3.59	-4.22	-5.53	-4.08	-5.80
bigbrain	2.96	0.78	1.32	1.31	0.64
smallbrain	2.54	0.44	2.09	2.47	1.46
no-feedback	1.94	0.64	1.58	2.20	1.18
no-agent	-1.61	-2.18	-2.75	-2.68	-4.29
3sides	3.33	0.64	2.56	3.13	1.59
w=a	0.13	-0.11	-0.45	-0.85	-2.36
no-penalty[a]	3.75 [-3.88]	2.83	1.95	0.94	0.31
blocked/no-penalty[a]	3.08 [-0.11]	0.61	1.44	1.74	0.75
blocked[a]	2.87 [2.88]	0.75	1.83	1.61	0.93

[a] The condition was tested in the environment with the interaction parameters they evolved in. The corresponding values of R (evaluated in the *Original* environment under standard settings: active penalty, blocking disabled) are added in parenthesis.

Table 3.8: *The p-values of the Mann-Whitney-U Tests for the average mean of the evolved fitness score $\langle EF \rangle$ per condition. The p-value for the preceded Kruskal-Wallis-Test is 0.000.*

/	**0.50**	**random**	**1.00**	**0.75**	**0.25**	**single**	**bigbrain**	**smallbrain**	**no-feedback**	**no-agent**	**3sides**	**w=a**	**no-penalty**	**blocked/no-penalty**
random	0.017													
1.00	0.000	0.000												
0.75	0.000	0.000	0.000											
0.25	0.000	0.012	0.000	0.000										
single	0.000	0.000	0.000	0.000	0.000									
bigbrain	0.453	0.204	0.000	0.000	0.071	0.000								
smallbrain	0.000	0.000	0.009	0.334	0.000	0.000	0.001							
no-feedback	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
no-agent	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
3sides	0.000	0.001	0.000	0.000	0.007	0.087	0.001	0.000	0.000	0.000				
w=a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.112	0.000			
no-penalty	0.000	0.000	0.000	0.000	0.000	0.344	0.000	0.000	0.000	0.000	0.017	0.000		
blocked/no-penalty	0.004	0.208	0.000	0.000	0.068	0.000	0.098	0.000	0.000	0.000	0.001	0.000	0.000	
blocked	0.435	0.048	0.000	0.000	0.001	0.000	0.372	0.000	0.000	0.000	0.000	0.000	0.000	0.012

Table 3.9: *The p-values of the Mann-Whitney-U Tests for the average brain complexity $\langle \phi_{Max} \rangle$ per evolutionary setup. The results of brain complexity calculations for $G_{bigbrain}$ are not available (NA) due to the computational complexity of the calculations. The p-value for the preceded Kruskal-Wallis-Test is* 0.000.

/	**0.50**	**random**	**1.00**	**0.75**	**0.25**	**single**	**bigbrain**	**smallbrain**	**no-feedback**	**no-agent**	**3sides**	**w=a**	**no-penalty**	**blocked/no-penalty**
random	0.409													
1.00	0.037	0.099												
0.75	0.023	0.041	0.370											
0.25	0.111	0.162	0.289	0.177										
single	0.000	0.000	0.062	0.177	0.010									
bigbrain	NA	NA	NA	NA	NA	NA								
smallbrain	0.000	0.002	0.080	0.191	0.029	0.496	NA							
no-feedback	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000						
no-agent	0.002	0.007	0.124	0.259	0.049	0.389	NA	0.421	0.000					
3sides	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.092	0.000				
w=a	0.000	0.000	0.000	0.001	0.000	0.001	NA	0.008	0.009	0.008	0.000			
no-penalty	0.000	0.000	0.008	0.023	0.001	0.105	NA	0.154	0.000	0.138	0.000	0.057		
blocked/no-penalty	0.196	0.375	0.134	0.063	0.331	0.001	NA	0.003	0.000	0.011	0.000	0.000	0.000	
blocked	0.395	0.494	0.079	0.026	0.145	0.000	NA	0.002	0.000	0.006	0.000	0.000	0.000	0.337

Table 3.10: *The p-values of the Mann-Whitney-U Tests for the average number of concepts in the set of elements with ϕ_{Max}, according to IIT, per evolutionary setup. The results of brain complexity calculations for $G_{bigbrain}$ are not available (NA) due to the computational complexity of the calculations. The p-value for the preceded Kruskal-Wallis-Test is* 0.000.

/	**0.50**	**random**	**1.00**	**0.75**	**0.25**	**single**	**bigbrain**	**smallbrain**	**no-feedback**	**no-agent**	**3sides**	**w=a**	**no-penalty**	**blocked/no-penalty**
random	0.372													
1.00	0.246	0.381												
0.75	0.070	0.120	0.272											
0.25	0.265	0.339	0.465	0.239										
single	0.007	0.014	0.083	0.221	0.051									
bigbrain	NA	NA	NA	NA	NA	NA								
smallbrain	0.000	0.000	0.010	0.019	0.002	0.044	NA							
no-feedback	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000						
no-agent	0.004	0.007	0.040	0.116	0.022	0.249	NA	0.187	0.000					
3sides	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.000	0.059	0.000				
w=a	0.000	0.000	0.000	0.000	0.000	0.000	NA	0.044	0.001	0.019	0.000			
no-penalty	0.000	0.000	0.002	0.011	0.001	0.042	NA	0.462	0.000	0.193	0.000	0.038		
blocked/ no-penalty	0.185	0.164	0.084	0.015	0.093	0.000	NA	0.000	0.000	0.000	0.000	0.000	0.000	
blocked	0.418	0.430	0.300	0.142	0.350	0.025	NA	0.001	0.000	0.013	0.000	0.000	0.000	0.204

Table 3.11: *The p-values of the Mann-Whitney-U Tests for the average reliability score $\langle R \rangle$ per condition. The p-value for the preceded Kruskal-Wallis-Test is* 0.000.

/	**0.50**	**random**	**1.00**	**0.75**	**0.25**	**Single**	**bigbrain**	**smallbrain**	**no-feedback**	**no-agent**	**3sides**	**w=a**	**no-penalty**	**blocked/no-penalty**
random	0.023													
1.00	0.000	0.000												
0.75	0.001	0.000	0.260											
0.25	0.198	0.002	0.003	0.008										
single	0.000	0.000	0.000	0.000	0.000									
bigbrain	0.450	0.077	0.001	0.003	0.210	0.000								
smallbrain	0.000	0.000	0.206	0.036	0.001	0.000	0.001							
no-feedback	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
no-agent	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000					
3sides	0.000	0.001	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000				
w=a	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
no-penalty	0.000	0.000	0.000	0.000	0.000	0.123	0.000	0.000	0.000	0.000	0.000	0.000		
blocked/no-penalty	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.041	0.000	
blocked	0.438	0.031	0.000	0.000	0.156	0.000	0.473	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3.12: *Spearman's correlation coefficient ρ between the evolved fitness EF and brain complexity ϕ_{Max} per condition (G_i). Note that weak or non-significant correlation coefficients may be due to small variance in the final evolved fitness values in some conditions.*

EF	ϕ^{Max}		$\#Concepts(\phi^{Max})$	
	ρ	**p-value**	**ρ**	**p-value**
0.5	0.6318	0.0002	0.6597	0.0001
random	0.2300	0.2215	0.0360	0.8504
1.00	0.4588	0.0108	0.6178	0.0003
0.75	0.2024	0.2834	0.1585	0.4028
0.25	0.6045	0.0004	0.4908	0.0059
single	0.1532	0.4189	0.3471	0.0602
bigbrain	NA	NA	NA	NA
smallbrain	-0.4112	0.1011	-0.2983	0.2449
no-feedback	-0.1034	0.5936	-0.1034	0.5936
no-agent	0.0439	0.8212	0.0768	0.6921
3sides	0.2847	0.1273	0.2804	0.1333
w=a	-0.0985	0.6047	0.0430	0.8214
no-penalty	-0.1293	0.4960	-0.1067	0.5748
blocked/no-penalty	0.2598	0.1656	0.4160	0.0222
blocked	0.4626	0.0100	0.6275	0.0002

Chapter 4

Quantifying the collective mind in organizations

Summary

Inspired by philosophies about the human mind, the idea of a mind at the collective level has spurred research over decades. Surprisingly, there is still little evidence of such a phenomenon. The absence of a visible body housing the mind of an organization and the complexity of mental processes themselves make it hard to argue that the collective mind is more than a metaphor. My study enhances the conceptual framework of the collective mind by analyzing the physical substrates of the collective mind and showing how they need to be structured to give rise to mental states in the organization. The integrated information theory (IIT) of phenomenological experience provides the theoretical basis for this approach. It allows for the measuring of the collective mind's physical structure, which declares the concept as a physical construct. Further, IIT's algorithmic framework can investigate the dynamics between micro- and macro-levels in the organization or the greyscales (different quantities) of organizational mental states. This theoretical expansion allows for the argument that a collective mind in organizations is a representation of organizational memory and can relate to organizational performance and decision-making.

Keywords: Collective mind; organizational cognition; organizational knowledge-structures; integrated information theory;

Co-Authors:	No co-authors.
Current-Status:	*Working Paper*
Paper History:	Accepted to AOM MOC Cognition in the Rough Workshop, August 2019, Boston, USA. Presented in the PhD Workshop at EGOS Colloquium 2019, Edinburgh, Scotland. Accepted to the Frontiers in MOC Conference, May 2020, Singapore. Accepted to the EURAM Annual Conference, December 2020, Dublin.
Acknowledgements:	I gratefully acknowledge the informal advice and support of Ravi S. Kudesia, Claus Rerup, Timothy Vogus, and Denis Grégoire.

"Where does the mind stop and the rest of the world begin?"
(Clark & Chalmers 1998, p.10)

4.1 Introduction

This work offers an expanded conceptual framework, as outlined by Shapira (2011), for the idea of a collective mind in organizations. The expansion involves the integration of studies about phenomenological experience (also known as consciousness), which makes the perspective on the collective mind more transparent, works up misconceptions, and offers methodological advances.

Small fish organize themselves into schools to increase their chances of survival. While swimming together, collective behavior emerges and the whole fish school seems to be one complex entity. Still, the fish only act to elementary rules, such as `'keep swimming'` and `'stay close to your neighbor'` (Garnier et al. 2007). In folk psychology, many believe that the human mind emerges from the controlled interactions of billions of neurons (Arico et al. 2011), similar to the fish school performing its collective behavior. If this emergence can give rise to the human mind, could it also give rise to a mind in fish schools (Bshary et al. 2014, Garnier et al. 2007, Will 2016)? Thus, this article examines whether collective behavior can give rise to a mind in organizations of mankind or not.

When theorizing about a supra-individual mind, the central assumption is that the physical substrates of a collective mind would be social, interrelating behavior (Sandelands & Stablein 1987, Durkheim 1895). If a neuron is a substrate of the human mind, social behavior would be a substrate of the collective mind. Indeed, the idea of a collective mind was already discussed by Aristotle and Plato (Voegelin 2000) and started to flourish in the nineteenth century when Le Bon (1896) and Freud (1922) discussed the psychology of the masses. In organizational studies, scholars have worked with the concept of a collective mind since the second half of the last century (Sandelands & Stablein 1987, Weick & Roberts 1993, Walsh & Ungson 1991).

I argue that revising the theory of a collective mind in organizations is necessary in order to promote empirical studies in that field. Repeatedly, the claims for the existence of a collective mind silenced after some time and became labeled as a metaphor (Jones 1995, Weick 1989). The key issue that prevents the investigation is that the collective mind is currently not observable, nor does it have a body. On the contrary, clinical studies can track brain activity in organisms and relate them to human emotions and behavior. For empirical studies, scholars have crafted implications and constraints for organizations with mental states at a collective level and have made them testable by conducting observations on the individual level (e.g., Vogus & Sutcliffe 2007, Sutcliffe et al. 2016, Oliver et al. 2017). Those works helped to validate and refine the theory and therefore support arguments that there is an objective phenomenon behind the idea of a collective mind.

In this work, I stress the position of viewing the organizational mind as equivalent to the collective mind. The collective mind would not only be shared or distributed between the individual minds of an organization (Hutchins 1995, March 1996) but would be clearly distinguishable from the individual mind (Gordon & Theiner 2015, Sandelands & Stablein 1987, Weick & Roberts 1993). This position allows for the validation of the theory and relevance of a collective mind in organizations and helps to find a method to investigate its structure.

The quality of a collective mind is determined by comparing organizational behavior with corresponding human mental states. Organizational mindfulness is probably the closest related and describes how organizations can stay reliable in uncertain environments (Weick et al. 2008, Sutcliffe et al. 2016). Initially, the theory was studied in High-Reliability Organizations (HROs) (e.g., wildfire firefighter or military units), where scientists try to measure the five hallmarks for organizational mindfulness in organizations (Weick & Roberts 1993, Fraher et al. 2017): Preoccupation with failure, reluctance to simplify interpretations, sensitivity to operations, commitment to resilience, and underspecification of structures (Weick et al. 2008). For instance, a military unit could barley prepare themselves for any possible future combat situation but could develop skills to avoid failure in uncertain environments. The soldiers would not simplify situations or small incidents that could accumulate and risk the whole mission. Detailed planning and continuous reporting, maintain an overall big picture of the whole mission,

to adapt their behavior when appropriate. In the case of situations that compromise the mission, the unit would try to save the mission, even if they would need to break up plans or hierarchy temporarily. Prior literature indicates cases in which organizations failed due to a lack in these processes or in which there is an improvement of organizations' performance by promoting the hallmarks in organizations (Aversa et al. 2018, Carlo et al. 2012, Rerup & Levinthal 2014).

It is inevitable to focus the observations on the individuals since the structure of the collective mind is unknown. But this method has problems answering the following questions (Sutcliffe et al. 2016, Aversa et al. 2018, Fraher et al. 2017, Carlo et al. 2012, Rerup 2004, Dernbecher & Beck 2017): First, prior literature complain that little is known about the dynamics between the organizational and the individual level within a mindful organization. Second, it should be possible to measure the quality and quantity of organizational mindfulness itself to find more about its structure and complexity. Finally, it is not clear how individual behavior contributes to organizational mindfulness in detail or if there is a causal relation between individual and organizational mindfulness.

The expansion of the underlying theory of a collective mind can help answer the open research questions. I elaborate on three steps inspired by the Domains-Interaction Model (Cornelissen 2005), explaining the usage of metaphors in organization theory. The model suggests to not only borrow distinct concepts from other fields, but also respect their semantic domain and how the metaphor itself would be constructed. I want to use recent cognitive theory to construct the collective mind:

The integrated information theory (IIT) of phenomenological experience (Tononi 2004, 2012, Oizumi et al. 2014) can be blended in to expand the collective mind. The link between both theories is the physical substrate of the mind. IIT provides an algorithmic framework to quantity the phenomenological experience of the mind, based on its physical substrates. First, I identify the common semantic structure between the collective mind and IIT. Current theory in organizational studies aligns to IIT's axioms and enables to connect the diverse concepts about the collective mind. Second, I develop the blend between the collective mind and IIT, which enables us to draw from insights made in IIT and to apply its measures. Finally, I craft implications and describe research methodology to be able to derive meaning for organizational theory and to enhance empirical research.

This study contributes to the literature in three ways. First, it offers arguments for the existence of a collective mind, based on interrelating behavior. Generally, I imply that the collective mind is a volatile phenomenon and hard to expect in routine situations. IIT supports that the collective mind would only exist as an integrated process between individuals' behavior. These findings confirm assumptions (Sandelands & Stablein 1987, Weick 1989, March 1962) based on recent neurobiological arguments. Second, aligning IIT enables to adapt algorithms used to measure the conscious mind in humans in a way to measure the collective mind, too. However, the quality of a collective mind is abstract and not imaginable for humanity (Nagel 1974) and if there was a collective mind, it would be extremely simple compared to the mind of human beings. Third, my findings contribute to the separation of the collective mind from organizational performance, which improves studying their relations. This would advance knowledge in studies about organizational mental states[1] and organizational memory.

This essay is structured as followed: I begin with a brief review about the collective mind in the organization. In particular, I analyze empirical studies to highlight conceptual issues, which can be resolved by reconceptualizing the collective mind. Then I form relations to the philosophy of mind and address such issues, while respecting the premises of the collective mind. Afterwards I align IIT's axioms with existing theory. Simple examples help to illustrate the application of the complex information-theoretical framework. In the latter, I make implications for related concepts in organizational cognition. Finally, I discuss the limits of the concept and how to conduct future research.

4.2 Theoretical foundations

Little is known with regard to how the collective mind of an organization could be captured. Researcher rely on measures at the individual level to craft arguments about its existence (Sutcliffe et al. 2016). This method limits

[1] E.g., organizational mindfulness.

our knowledge in that scope, which is why we draw on advances in research about phenomenology to enhance the concept of a collective mind. In past decades, neuroscientists have made great advances in understanding how the mind emerges within ourselves.[2] A kind of mystic frontier in that field is to explore human phenomenological experience by investigating its physical footprints (Koch et al. 2016, Baars 1988, Tegmark 2015). As already done before in organizational cognition, I borrow insights from neuroscience to expand upon the conceptual framework of a collective mind. I analyze works related to Sandelands & Stablein (1987) and to the ideas in Weick & Roberts (1993), that are early milestones in organizational cognition.[3]

4.2.1 Collective mind and organizational mindfulness

Generally, the collective mind in organizations is a concept to describe the representation of *supra-individual knowledge structures* (Walsh 1995). The physical substrates[4] of a collective mind are said to be social, interrelating behavior. The idea that the collective mind's basis is strongly interrelating behavior (Sandelands & Stablein 1987) refers to an area of cognitive science that attempts to describe the mind through neuronal activity.[5] Weick & Roberts (1993) continue those thoughts and relate them to the high-reliability of organizations that perform nearly error-free in highly dynamic and uncertain environments.

Scholars investigate specific mental states of organizations (e.g., *attention* or *awareness*) by assuming a collective mind or at least using it as a metaphor. This allows one to predict the functionality of a collective mind and to imply qualities that are similar to those of human mental states. It was necessary to move from discussions of a collective mind to specific mental states to make the concept observable, since the collective mind itself does not have a body, nor is its physical structure known (Wegner 1987).

I specify this interpretation using the concept of organizational mindfulness, as an example of a mental state in an organization. There is already a stack of empirical research and operationalizations on that topic (Black & McBride 2013, Gebauer 2017), and organizational mindfulness is also closely related to the concept of a collective mind in organizations (Weick & Roberts 1993, Sandelands & Stablein 1987). Briefly, the difference between the mind and mindfulness is that the mind would embed mental states like mindfulness as the quality of conscious experience (Brown & Cordon 2009). Recent works in that field use organizational mindfulness more as an anthropomorphism that points to specific organizational performance or behavior.

Organizational mindfulness explains why organizations stay reliable in very dynamic environments (Weick & Roberts 1993). It is used to study high-reliability organizations as they occur in military, firefighting, or healthcare. In such organizations, even a small mistake causes life-threatening damage, while the organizational members often face incidents that are close to catastrophes. Weick and colleagues argue that interrelating behavior is the substrate of the organizational mind, which produces a state of mindfulness in the organization. Prior research measure organizational mindfulness based on the predominance of five processes[6], the hallmarks for organizational mindfulness (Weick et al. 2008, Weick & Sutcliffe 2015):[7] (1) Preoccupation with failure, (2) reluctance to simplify interpretations, (3) sensitivity to operations, (4) commitment to resilience, and (5) underspecification of structures.

Sutcliffe et al. (2016) offer a general review of organizational mindfulness, mindfulness in organizations, and which methods for investigation are available. An important contribution of their review is the comparison of the different definitions of (organizational) mindfulness, which have turned out to be less precise than anticipated. This comparison and their calls for researching methods to investigate the structure of organizational mindfulness more scientifically supports the claim that we need to expand upon the concept of a collective mind.

I explain the difficulties of investigating organizational mindfulness by presenting three selected studies: First, Oliver et al. (2017) analyze a business simulation experiment to validate that mindful organizing is not only suitable

2 While in the eighties, computers were used to describe the human brain, nowadays knowledge about the human brain is used to advance computer science and artificial intelligence.

3 My discussion can also be related to the concepts of memory and knowledge structures in organizations (Wegner 1987, Walsh & Ungson 1991, Joseph & Gaba 2019, March 1996), but those should be a subject in the discussion section.

4 A physical substrate is the smallest objective building block of a mind (also related to micro-foundations).

5 See connectionism (Long 2005).

6 Recently additional processes were suggested by Fraher et al. (2017) and Oliver et al. (2017).

7 A more profound explanation of the processes is beyond the essay's topic. Please refer to the cited works for more details.

for reducing failures but also for increasing the overall performance of the organization (e.g., less accidents). They adapt the questionnaire-based safety organization scale (Vogus & Sutcliffe 2007) to measure mindful organizing. They find a positive correlation between team performance and mindful organizing. The methods used to identify mindful organizing can only measure the individual's perception of the social behavior in the team and do not reveal much about the actual behavioral structure of the collective mind and its communication processes.

Second, Vendelo & Rerup (2019) present an ethnography to investigate collective mindfulness in a regenerating organization.[8] They collect data in the crowd safety teams of an annual music festival and test the evidence of the hallmarks for organizational minding (Weick et al. 2008, Weick & Sutcliffe 2015). The theory suggests that organizational mindfulness is changing dynamically, which is already highlighted in Rerup (2004) and Vendelo & Rerup (2019). As a consequence, they investigate events that could cause a shift in the quality of organizational mindfulness. Although they had no particular focus on investigating the cognitive structure, they find evidence for different categories of organizational mindfulness and that it is present at different organizational levels.

Third, Fraher et al. (2017) present a study on the assessment of candidates for the U.S. Navy SEAL special forces. This study investigates both individual and collective mindfulness. Physical skills alone are not suitable predictors for excellent performance during the extreme admission test for becoming a SEAL member. That psychological and team abilities might play a central role was a motivator for investigating the mindful qualities of individuals and the troop as a whole. Their study finds that *comfort with uncertainty* is a further hallmark of organizational mindfulness. Concerning this analysis, it is essential that the authors clearly distinguish between individual and organizational mindfulness. In doing so, they find need for research in understanding the relationship between organizational and individual mindfulness.

Studies about organizational mindfulness are not always consistent in distinguishing the concept as a shared, distributed, joint, or collective mental state (e.g., awareness, attention, focus). I see a need to discuss the used semantic domain to find common ground. In this approach I claim that this ambiguity is also due to the limited knowledge about the collective mind and that we should not investigate the structure of a collective mind without a uniform definition of the conceptual framework.

In general, I see substantial research interest in the concept of organizational mindfulness, but barriers, like overcoming the strong focus on the individual level, impeding investigations at an organizational level. I believe that not knowing much about the structure of a collective mind is fundamental for such gaps. Moreover, it is often inconclusive how studies interpret the collective mind and organizational mindfulness (Sutcliffe et al. 2016, Dernbecher & Beck 2017), suggesting the need to expand the conceptual framework. In a similar vein, Walsh suggested that we should *"move beyond individual minds in our considerations of supra-individual knowledge structures"* (Walsh 1995, p. 311). I want to validate the collective mind as a supra-individual knowledge structure (Walsh 1995) using arguments borrowed by neuroscientific theories. The lowest common denominator for the alignment of both disciplines are interrelated mechanisms as the physical substrates of the mind.[9]

4.2.2 Conceptual origins

Drawing on the foundations of organizational mindfulness, we can identify its relation to the collective mind itself. While organizational mindfulness is mostly grounded in Weick & Roberts (1993), it relates to discussions on a more general level (Wegner 1987, Walsh & Ungson 1991, Sandelands & Stablein 1987). The discussions are mainly conducted with the assumption that interrelating and redundant social behavior lead to reliability and *smart* behavior within organizations. The origin of that assumption is in the works of Durkheim and his colleagues (Durkheim 1895, Halbwachs 1952, Fleck 1983).

Durkheim laid the foundation for projecting society as a distinct object and questioned what is needed to hold the society, as a set of individuals, together.[10] A modern society, where individuals differ very much between one another, is held together by high interdependency and requires social facts to produce this solidity. Social facts are

[8] Regenerating organizations are organizations that exist only temporarily.

[9] The human mind's physical substrate might be the neural network, which can be aligned to the network of social behavior within the organization.

[10] Also known as *solidity* (Durkheim 1895).

external to the individuals and override their egoism by society's shared myths, like the knowledge of the language, institutions, or law. He also defined the idea of *collective consciousness*, which should not be confused with the collective mind, as we use it in this work[11]. Durkheim's aim to move sociology into an objective science requiring society to treat and its social facts as distinct objects. The following statement summarizes this school of thought:

> *"That the content of social life cannot be explained by purely psychological factors, namely by states of the individual consciousness, seems to us to be as plain as can be. Indeed what collective representations express is the way in which the group thinks of itself in its relationships with the objects which affect it." (Durkheim 1895, p. 40)*

Durkheim's colleagues continued his thoughts and realized that the objective investigation of social facts in organizations might be challenging (Fleck 1983) and tried to theorize about the structure of a collective mind (Halbwachs 1952): While investigating the social system, the observer is usually also part of the society and would share similar social facts, which produces a significant bias, and the observer might not capture essential behavior and routines that could be substantial for collective behavior (Fleck 1983). An approach for explaining the gap between the individual and its social environment was delivered by Halbwachs (1952). He stressed that conceptions of the past are affected by the mental images we employ to solve present problems. Hence, memories are a reconstruction of the past in the light of the present. For example, some soldiers use popular media to recreate memories of their combat missions (Nguyen & Belk 2007, Walsh & Louvre 1988). In this manner, short-term behavior in organizations influences the retrieval of individual memory. This effect supports critiques about collecting observations at the individual level only, while ignoring the social system the subject is part of. Also, the importance of the social structure itself motivates me to put more focus on the investigation of the relations between individuals in order not to depend solely on the introspection of the subjects.

Especially issues with observations at the individual level might imply that mental states like organizational mindfulness are phenomena of aggregated individual behavior (Jones 1995, March & Simon 1958, Simon 1947). However, by reducing the idea of the collective mind exclusively to an aggregation of information processing between individuals, it would be merely an umbrella term for joint and individual mental states. Nevertheless, scholars had a phenomenon in mind that is irreducible to the individual (Wegner 1987, Weick 1976, Sandelands & Stablein 1987).

4.2.3 Representing the collective mind

For an organization of individuals to emerge as a collective mind, their behavior must be tightly coupled. That means that there might be redundant, highly-interrelated processes within the organization. The emerging structure also refers to *tightly coupled systems* (Weick 1976, Orton & Weick 1990):

> *"Tightly coupled systems are portrayed as having responsive components that do not act independently, whereas loosely coupled systems are portrayed as having independent components that do not act responsively." (Orton & Weick 1990, p. 205)*

The specific behavioral structure of a tightly coupled system is essential for their effectiveness. If looking at the tightly coupled system of a school of fish, as described at the beginning, the right structure can create positive effects for the organization. Hence, we would need to answer the question about how the physical substrates need to be structured to give rise to the collective mind. This question is not exclusive to our problem but can also be transferred to theories about the human mind: How do the neurons within our head need to be structured and activated to give rise to mental states (Tegmark 2015)? The cerebellum, for example, contains most neurons within our brain, but is not responsible for giving rise to the conscious mind (Koch 2018).

Before we examine the structure of the collective mind, we must first understand how we can imagine it as an independent object and that we should hide the individual mind for further consideration. A symptom of the

[11] Collective consciousness explains that a traditional society is held together by shared beliefs and similarities.

lack of knowledge about the specific structure of a collective mind is that we know little about how to differentiate between the levels of the collective mind and the individual mind. I describe the difference and how collective mental states can emerge from the individual mind by adapting a model that describes collective intentionality:[12]

1. *Individual mind:* Every individual has its mind and mental states, which are exclusive to the individual. The individual itself produces all conscious and unconscious mental states and its mind is bound to its body.

2. *Joint mind:* The aggregation of individual minds creates a joint mind. For instance, when hunters had to hunt a mammoth, they had joint intentions, but could not reach the goal when acting exclusively alone. Hence, they shared concepts like the motivation (*we need food*) or skills for a successful hunt (*we have to corner the animal*). The ideas of a shared, distributed, and extended mind (Hutchins 1995, 1989) operate at this level. Goals, plans, and actions are shared socially, even though there is individual differentiation.

3. *Collective mind:* The aggregation of joint minds is the foundation of a collective mind. Social facts can be the scaffolds for the group mind to coordinate the individual's behavior within the collective (Tomasello 2014).

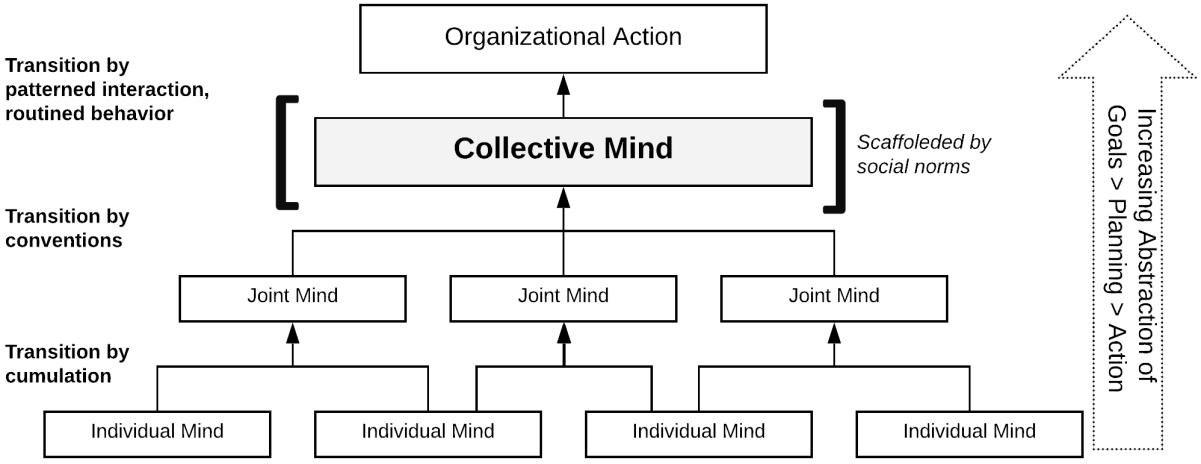


Figure 4.1: *The transition from individual mind to joint mind and collective mind, adapted from Theiner & O'Connor (2010).*

The level of abstraction (e.g., of goals, plans, and actions) increases when moving from the individual to a collective mind. For instance, on a flight carrier, each crew member has a distinct duty, which extends from the *cleaning of the flight deck* to the *steering of the ship*. Only with the combination of all actions can the carrier be safely operated as an airport on the sea (Weick & Roberts 1993). This perspective contains two essential arguments for a collective mind in organizations. First, the joint mind or an aggregation of it cannot be a synonym for the collective mind. Second, the structure of a collective mind doesn't have infinite possibilities. It is constrained by our social facts, from language to law.

However, we cannot observe the collective mind as we would the activity of neurobiological networks in organisms to identify the structure of their mind (*outside-in*, e.g., by electroencephalography (EEG)). Even if this were possible, there is an explanation gap between the observed activity and the subjective mental states (Koch et al. 2016).[13] The only possibility would be to investigate the collective mind from its physical substrates, the social behavior within the organization (*inside-out*), which is the approach in this study. To do so, we need to apply neuroscientific theory, which explains how social behavior can construct a conscious mind.

[12] See Figure 4.1, based on Tomasello (2014), Gordon & Theiner (2015).
[13] Few is known about how neural activity can explain the quality of subjective perceptions.

4.3 Quantifying the collective mind

This section shows that the theory of a collective mind can be improved by returning to its conceptual roots and linking them to recent advances in cognitive science. Generally, there is already a blending of the theory of mind with fields of neuroscience or even artificial intelligence (e.g., Nelson 2002, Rosenthal 2005, Searle 2000). The abstract character of these discussions makes it possible to gain insights for organizational cognition, too. I build on the foundations of the theory of mind and neuroscientific studies to work on problems in the concept of a collective mind, taking into account its premises. The link forms the translation of the Integrated Information Theory (IIT) (Tononi 2004, Oizumi et al. 2014). In particular, I show how its axioms coincide with the idea of a collective mind in organizations.

4.3.1 The concept of a mind revisited

Before I blend in IIT, I review the definition and function of the mind itself and how it relates to the physical world. Many use the term *mind* in everyday language, without being confident about its precise definition. Traditionally it is the subject of philosophers to explain cultural terms like *mind, consciousness, thought*, or *phenomenology*.[14] However, when we apply such concepts at the organizational level, we need a uniform definition of the underlying concept of the human mind, too. Confusion and obscurities are inevitable if we only rely on general knowledge. Even when the concept of a collective mind differs from that of a human equivalent, outlining the differences would be helpful guidance for readers.

A short analogy should help to categorize the concepts (Block 1995, Searle 1980): If we were computers, the brain could be the hardware implementing the mind as software. The software would enable processes like thinking, feeling, and controlling our awareness. Without the software, the hardware alone could not perform cognitive processes. We know that there are processes that we are aware of[15], and others that are out of our direct control[16]. The subject, which witnesses the conscious mind, can be called phenomenological experience – Descartes tried to describe it by his famous quote: *"I think. Therefore I am"*. Translated into modern language, it would mean something like *I am experiencing the current moment, which is proof of my physical existence,* (Oizumi et al. 2014). Hence, a mind implies the existence of a physical representation (the brain) but also needs an intrinsic observer (the conscious mind, also known as phenomenological experience or consciousness). It is difficult to distinguish between the brain, the mind, and the phenomenological experience. Still, we do not know if our subjective world is bound to our body or not (Chalmers 1997). Further, psychologists and neurobiologists show that not everything mental is conscious, but anything conscious is mental (Searle 1991).

It is crucial to include phenomenological experience in the discussion of a collective mind. Anthropomorphisms in organizational studies (e.g., organizational mindfulness) refer to mental states that we experience consciously. Since the conscious experience is a fundamental property of our mind, we first need to show how a collective mind can give rise to phenomenological experience before ascribing qualities like mindfulness to it. Discussing phenomenological experience at an organizational level can serve as an essential tool for translating mental states. For example, if I share the same strawberry ice cream with my friend, I would imply that my friend has a similar experience of the ice cream's taste. In that manner, if we attribute mental states to organizations, we imply that organizations have the qualities associated with those mental states. For instance, if a mindful organization had a high quality of situational awareness, we would subjectively compare it with the quality of our own awareness.

The concept of *intentionality*, discussed by Husserl (Schmitt 1959), allows one to distinguish the mind from its *experiencing* entity. Husserl aimed to make sense of the relation between the perceived experienced, *the inner world*, and the environment of the subject, *the outer world*. His most important contribution was to enable a logical analysis of something subjective, *how psychological phenomena relate to something objective*: The mind has phenomenological experience about an object, while the object itself has relations to all other objects. For instance,

[14] See also the discussions about the theory of (Voegelin 2000, Schmitt 1959).
[15] E.g., our vision as part of the conscious sensory perception.
[16] E.g., the regulation of our heartbeat.

comparing tulips and roses, they have similarities and differences, and we might associate special memories including the flowers, e.g., we could remember tulips at a funeral and roses at a wedding. Notable is that intentionality contains phenomenological reduction. Phenomenological reduction describes minding without knowing or having conceptions about an object, respectively, a conscious perception of something without any judgment of the situation. This concept distinguishes experience from pure functional cognition. Searle offers a suitable definition of intentionality:

> *"'Intentionality' is a word that philosophers use to describe that feature of the mind by which it is directed at, or about, or of, or concerns, objects and states of affairs in the world." (Searle 2005)*

As I argued before, I do not suggest to observe the collective mind's activity *outside-in*. To investigate it *inside-out* we would need to investigate the physical representation. For the collective mind, it makes only sense to consider the mind as a physical phenomenon, since theory already assumes individual behavior as the physical substrates of a collective mind. This argument translates to the view that matter, structured correctly, gives rise to phenomenological experience (Tegmark 2015).

Figure 4.2 illustrates an ontology about the semantic domain the collective mind is embedded in: Integrated organizational elements that act in and with an environment define the objective world. Dependent on how the individuals are structured and how they interrelate, they might form an organizational brain. If the brain has a particular structure and activity, the mind emerges as a subjective phenomenon. The mind itself creates abstract perceptions of the external environment. Depending on the complexity of the mind, it can embed mental states that could be experienced by itself, too. The collective mind would produce organizational action, which influences the organizational elements and therefore creates an interrelating complex.

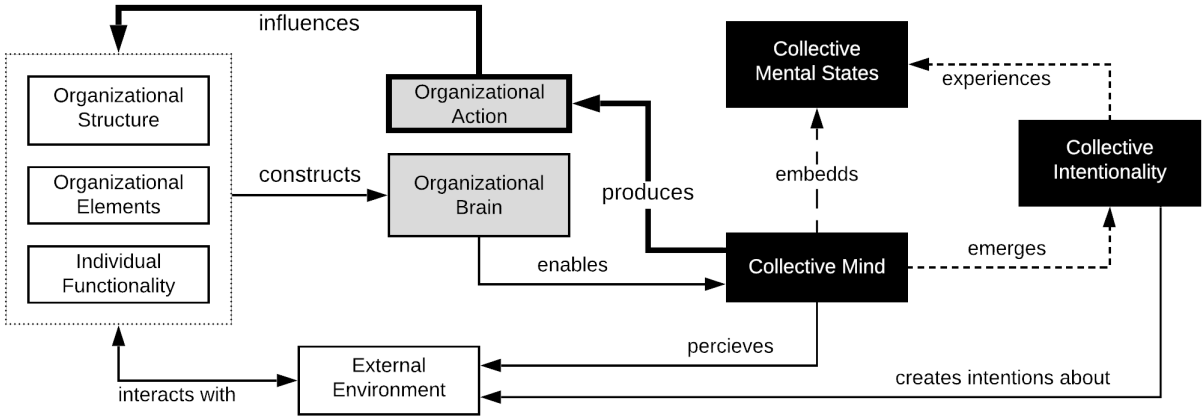


Figure 4.2: *The conceptual framework, which explains the relationship between the collective mind and its physical substrates.*

4.3.2 The structure of the collective mind

Until now, I have made sense of the concepts and how they interrelate with each other. Next, I investigate the physical structure of the collective mind. To make sense of phenomenological experience at a collective level, it needs a theory that can explain it on a physical basis. Inspired by neural correlates of consciousness (NCCs), Tononi et al. developed IIT (Tononi 2004, Tononi & Koch 2008, Oizumi et al. 2014) with the intention to assess the state of consciousness in apparently unconscious patients, e.g., coma patients or test persons during sleep experiments. IIT provides an algorithmic framework to quantify the conscious experience of the mind based on its physical substrates. The framework is known for taking the physical existence of conscious experience as a fundamental axiom. Based on the axioms, the theory postulates its physical aspects. By doing so, it is not only possible to apply IIT to humans and other organisms, but also organizations.

The approach that IIT is following is in line with other cognitive or neurobiological theories (Crick & Koch 2003, Edelman 2006, Dehaene et al. 2006, Baars 1988) on fundamental assumptions: First, distributed, highly integrated, and interactive networks give rise to conscious experience. Second, conscious experience and mental states overlap, but not in the same way, e.g., there is attention that is experienced and simultaneously sub-conscious attention. Finally, IIT respects that a cognitive system has unconscious sub-systems whose investigation can provide new insights, too.

IIT is a suitable framework for explaining phenomenology in a collective mind and to enhance the concept itself. IIT's interrelations with other cognitive theories support the decision to use it for explaining the structure of a collective mind. Unlike conventional cognitive theories, IIT does not attempt to answer how an organism could embed the phenomenological experience (*outside-in*). It is the other way around, since IIT defines phenomenological experience as a fundamental truth and postulates rules for a physical system that give rise to conscious mental states (*inside-out*). The five axioms for phenomenal experience in IIT are *existence, composition, information, integration,* and *exclusion*. See Table 4.1 for an overview of how they are defined and how they align with the collective mind. A detailed explanation will be the subject of the next section.

Table 4.1: *Aligning IIT's axioms and the premises for the collective mind. The columns show axioms as undeniable truths about conscious experience, postulates as implications for the physical representation of the axioms, and the description how both link to the concept of the collective mind.*

Axiom	Postulate	Link to the collective mind
Existence: The experience of mental states is real, like Descartes describes it in his quote *"I think therefore I am"*.	Substrates of conscious experience exist physically and can be in a specific state.	If there is a collective mind, it has to have physical substrates. In an organization it is individual behavior that emerges the collective mind.
Composition: Mental states are structured. Each moment contain different aspects and variations, e.g., a cup of coffee is composed by sensing liquid, heat and taste.	The mechanisms of the mind can be combined into higher-order mechanisms with joint functionality.	Individual behavior can be joined into higher-order behaviors to fulfill organizational goals, e.g., joint action of construction workers to build a skyscraper.
Information: Experiencing mental states contain information and is different from other experiences, e.g., a cup of coffee differs from a cup of tea by taste and color.	The physical substrates can contribute to the current experience only if they make a difference within to it. This is only possible if the substrates interrelate with other mechanisms.	Only if individuals react to other's behavior and if the behavior of others is influenced by the individual it can be part of the collective mind, e.g., an idle construction worker would not be part of the joint action of building a skyscraper.
Integration: Phenomenological experience cannot be divided into independent components, e.g., when experiencing a sip of coffee you cannot separate the experience of the taste from feeling the liquid's temperature.	Only if the information, the mechanisms generate, constrain the future and past states of other mechanism, they could be part of the current experience.	Only complex, interrelating social behavior might give rise to the collective mind, while the intrinsic information of the mental state is irreducible to individual behavior.
Exclusion: The conscious mind is unique, constrained by space and time. It is not possible to experience different levels of mental states simultaneously – *"each experience differs in its particular way from other possible experiences"* (Oizumi et al. 2014, p. 2).	A certain physical mechanism can only contribute to the most complex sub-system, which stores the highest integrated information.	Simultaneous individual behavior could be clustered in various ways, but only the most complex combination of social behavior might contribute the the collective mind.

IIT's postulates are the foundations for its algorithmic framework. This framework provides methods to determine the causal power for phenomenological experience in a physical system. The causal power describes the contribution of one or more physical substrates to the whole phenomenon of a conscious mind. The measures of Φ (*"big phi"*) determine the quality and quantity of integrated information in a physical system. The higher the value of Φ is, the more complex the mind would be. Thus, an organization might have conscious mental states only if its value of Φ is greater zero. In short, IIT measures how much information the system has intrinsically by quantifying the loss of information when partitioning the system in all possible subsystems. In an organization, we can measure how much an individual's behavior contributes to the collective mind and we can determine whether the collective mind would break down (loss of all integrated information) when specific individuals drop out (partitioning of the system).

4.3.3 Measuring the collective mind

The five axioms for phenomenological experience will be the foundation of the alignment process (Cornelissen 2005). I introduce each axiom's meaning, how it postulates to a physical system, and how it translates to the concept of a collective mind, similarly as in Table 4.1.

Existence

The existence axiom links to the definition of individual behavior as the physical substrate of a collective mind. Physical substrates are the elemental mechanisms that are causal for intrinsic information within the collective mind as a whole. Formally, the elemental mechanisms $m(i, b) \in M$ of a system are an individual $i \in I$ who has a behavior $b \in B$. The repertoire of all elemental mechanisms would be $M = I \times B$. Each mechanism m has a discrete state $s = S(m(i, b), t) \in \{0, 1\}$ at a specific moment $t \in T$. This definition could lead to $|I|\times|B|$ different elemental mechanisms. An active mechanism ($s = 1$) represents an individual in a specific action, e.g., the pilot talks into the radio. A mechanism can also be disabled ($s = 0$), like the pilot who forgets to check the flight altitude.

Composition

The composition axiom links to the aspect that individuals join to achieve higher-level goals. Elemental mechanisms can form a higher-level mechanism according to the composition axiom, e.g., when we see a car, we see not only its shape but also its color and cannot separate both in our experience. In an organization, the combination of a set of individuals' behaviors can combine to a higher-level mechanism. For example, the process of landing a jet on a flight deck, where many crew members work together (Weick & Roberts 1993) or the operation of a jet cockpit (Hutchins 1995), shows that a composition of individual behavior is necessary to achieve higher-level goals.

Figure 4.3 shows an example of a simple organization of three individuals. Each has three behaviors available, where only one behavior is active (black circles). There are two individuals (grey), who are not integrated in the organization. All three individuals in the organization are connected by interrelating actions. Hypothetically, the three central individuals can be in charge of joint action, like jointly changing the wheel of a race car. This example displays the implementation of IIT's existence and composition axiom, which maps the approach to describe the collective mind through connectionism (Sandelands & Stablein 1987).

Information

Collective mental states contain information, just as phenomenological experience does. The smallest entity of information in a collective mind is social interrelating. In IIT it is essential to not view information as sender-receiver information, as Shannon (1948) did, but as intrinsic information, represented by the organization's social behavior itself (Tononi et al. 2016). Since the behavior of individuals is continuously changing, there is a difference

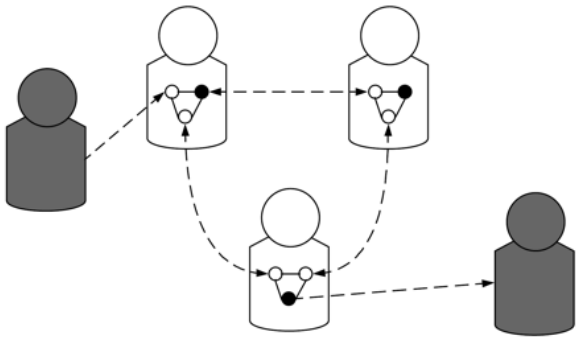


Figure 4.3: *Example, to visualize the physical substrates of a collective mind in an organization. The system shows three individuals having a repertoire of three actions each (small circles within an individual). The actions can be active (black circle) or inactive (white circle). The individuals are not connected physically but indirectly through their interrelating behavior.*

between all perceived moments. A system of behaviors is a dynamic process and will differ from one moment to another. The causal role of individual behavior to the intrinsic information of the organization would be determined by how much the current behavior constrains the future or is constrained by the past – the mechanisms have to make *"...differences that make a difference"* (Oizumi et al. 2014, p. 3). For instance, during a race the mechanic of a NASCAR team might contribute more to the organization's performance than its press spokesman.

We can measure the information generated by an individual's behavior by calculating the cause-effect-information (cei) as the difference between causes/effects, including the particular mechanism versus excluding it. For example, if considering a sports team and its coach: If the coach's command did not affect the team, its cei would be zero, and, vice versa, if the team's behavior were not affected by the coach's commands, it would have zero cei, too. In a sports team, where coach and players are highly integrated, they would have a $cei > 0$.

It is possible to calculate the integrated information φ (*"small phi"*) of an elemental or joint mechanism by considering its cei and the irreducibility of underlying mechanisms by partitioning the system and testing what difference it makes if the mechanism is missing. For example, if an individual's behavior is not causal to the organizational system, it would not contribute to the collective mind. Mechanisms with $\varphi > 0$ are so-called *concepts*. Only concepts can contribute to the conscious mind. To determine the quality of experience – *what it feels to be in that experience* – it is necessary to calculate the integrated conceptual information φ of all possible mechanisms of the system (elemental and higher-order). If $\varphi > 0$, the mechanism is a *complex*. The constellation of all concepts with integrated information φ form a complex and represents the *quale* of conscious experiences (Oizumi et al. 2014).[17]

Integration

For a collective mind, it would be essential that social behavior is strongly interrelating. This links to the integration axiom in IIT. If an individual acts randomly, without any respect for its effects, it could not contribute to the collective mind. If, however, the individual's behavior has a high impact on the behavior of the whole organization, it would have a high power to change integrated information φ. Further, the organizational behavior that the collective mind produces would not be (fully) reducible to its individuals. For instance, the landing of an jet plane on an aircraft carrier involves various sub-processes carried out by many individuals, and the landing itself cannot be reduced to individual persons (Weick & Roberts 1993).

If a mechanism, or a combination of mechanisms, is strongly integrated, it can create integrated information φ. Only systems with φ can represent a conscious mental state. Hence, individual behavior can only contribute to the collective mind if it influences others' behavior and vice versa. An individual has to perform actions that are constrained by other mechanisms and will constrain mechanisms in the future. The cause-effect repertoire of

[17] I avoid a deeper description of the measures, since the algorithmic framework as presented in Oizumi et al. (2014) is non-trivial.

a mechanism represents all possible past and future states of the system, constrained by the current state of the mechanism. IIT not only measures the contribution of a single individual to integrated information but also considers all possible higher-level mechanisms. Further, we can determine how much joint behavior would influence the collective mind itself, since not only individual behaviors are mechanisms, but also joint behavior can be seen as higher-level mechanisms.

Exclusion

Not every individual in an organization contributes to a collective mind similarly. Hence, assuming that the whole crew of an aircraft carrier might contribute to the collective mind in a similar manner or at all is incorrect. If we were to observe social behavior on the bridge of the ship, we would need to respect that the behavior between two navigation officers might be more integrated (high φ) than the behavior of the whole bridge crew (low φ). To determine which cluster of integrated social behavior represents the collective mind, we would need to postulate IIT's exclusion axiom.

The exclusion axiom describes that only the most integrated concepts of a system can contribute to conscious mental states. For a collective mind, a concept is a set of individual behavior that has a causal role for social behavior. The behavior of the concept is further irreducible to its parts. Taking the above example: If the integrated information φ between the two navigation officers were higher than the integrated information φ of the whole bridge crew, only the two officers would represent the collective mind of the whole bridge crew, which is indicated as Φ.

Further, it is possible that the whole crew gives rise to multiple concepts (groups with integrated information φ). Such can be small local units that are strongly interrelated. For phenomenological experience, only the concept with the highest integrated information (Φ^{Max}) would take over the role of our consciousness. Otherwise, it would be possible to have multiple personalities within one body simultaneously. This concept is causal for the current emotion – how it feels to be in that moment (phenomenological reduction). For now, that extension is not relevant when investigating the collective mind, since each subsystem with integrated information would be of interest.

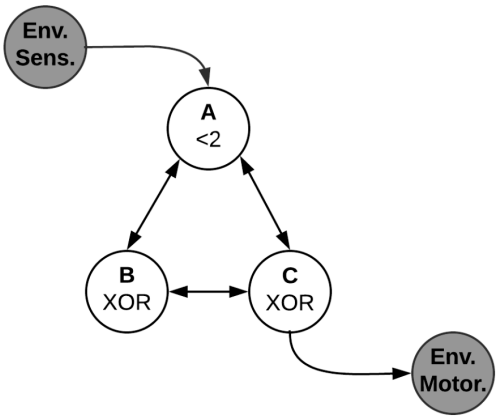
4.3.4 An illustrative demonstration

A minimalistic demonstration should help to explain how to apply IIT's algorithms in an organizational system. Figure 4.4 shows a system with five elemental mechanisms. There are two environmental mechanisms (grey) and three mechanisms within an integrated system. Mechanisms A, B and C represent individual behavior interacting, while a single individual only has one behavior available.

Hypothetically, A could have the role of a group leader who turns active if the activity in the system is too low. B (or C) only becomes active if C (or B) is inactive and A is active or if A is inactive, and C (B) is active. This scenario can represent load balancing in a system: Imagine two construction worker and their supervisor while shoveling soil into a truck. They alternate each couple of minutes, and only if they are both inactive does the supervisor have to push them. The essential factor in enabling integrated information is that all elemental mechanisms in the subsystem ABC give feedback to each other (consider the connections between mechanisms). In the current state, A, B, and C are not active, and the state of the subsystem ABC is $(0, 0, 0)$.

As displayed in Figure 4.4(a), it can be that logical rules activate the individual behavior. For instance, behavior B (XOR is the abbreviation for exclusive OR) would only be active in the future (t_1) if exclusively behavior A or behavior B is active before (t_0). A becomes active if two other connected mechanisms are active. B and C shift to an active state if precisely one connected mechanism is active. The environmental mechanisms are not integrated into the system and can, therefore, not contribute to a collective mind. They only visualize the possibility of information exchange with the external environment.

This formal notation is required to be able to apply IIT's algorithms. The *transition probability matrix (TPM)* displays the probability of future states in the system, depending on the system's state at the current moment. The TPM results from gathering all possible system states in preceding observations. In our demonstration, logic rules



t_0 \ t_1 ABC	**000**	001	010	011	100	101	110	111
000	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
001	**0**	0	0	0	0	0	0	1
010	**0**	1	0	0	0	0	0	0
011	**0**	0	1	0	0	0	0	0
100	**0**	0	1	0	0	0	0	0
101	**0**	1	0	0	0	0	0	0
110	**0**	0	0	1	0	0	0	0
111	**1**	0	0	0	0	0	0	0

(a) **(b)**

Figure 4.4: *Logic gates are a method to model social behavior: (a) A sub-system ABC having three elemental mechanisms and two background mechanisms (grey) in the state $ABC = (0, 0, 0)$. The sub-system exchanges information with the environment through the background mechanisms. (b) The system in (a) produces a transition probability matrix (TPM) by their specific function (here logic gates). The TPM displays the probability for future states at t_1 constrained by the current state at t_0.*

constrain the system's states. According to the TPM of the above example (see Figure 4.4(b)), only six of the eight possible states are valid (two states are logically impossible). IIT uses information-theoretic measures to calculate the integrated information by considering the TPM and the connectivity of the system. In the current example, the maximal integrated conceptual information Φ^{Max} is 3.23. The main complex, which has the maximal value of Φ is the joint mechanism ABC. ABC has six subsets that form a concept ($\Phi > 0$) and therefore contribute to the overall experience of the subsystem. Only mechanism A cannot form a concept.

The above example investigates only one state $ABC = (0, 0, 0)$, which has the highest Φ^{Max} over all states.[18] Since this state has the highest integrated information we know that it maximally constrains all possible past and future states. However, also considering the average values over all states can highlight the characteristics of an organizational system. The mean integrated information of the main complex is $\langle \Phi^{Max} \rangle = 2.08$. This value is about one third lower than the maximum. Considering Φ^{Max} over multiple states and investigating its distribution can be an indicator of the collective mind's robustness. The minimum in this subsystem, would be $\Phi^{Max} = 1.00$ in the states $ABC = (0, 1, 0)$ and $ABC = (0, 0, 1)$. If high Φ^{Max} is dominant in only a few states, the collective mind would be a very volatile phenomenon. If Φ^{Max} is uniformly high across many states, it means that the collective mind appears as a constant trait in the system.

Table 4.2: *Information theoretic analysis of the demonstrated organizational system. Each row represents the IIT-related values per state. All values were calculated with the programming package PyPhi (Mayner et al. 2018).*

ABC	Φ^{Max}	$\Phi^{Max}_{Elements}$	$\Phi^{Max}_{Concepts}$	$\sum \varphi^{Max}$	$\sum \varphi^{Max}_{Concepts}$
(0,0,0)	3.23	3	6	2.17	6
(0,0,1)	1.00	2	2	1.17	4
(0,1,0)	1.00	2	2	1.17	4
(0,1,1)	2.10	3	5	1.77	5
(1,0,0)	2.94	3	5	2.00	5
(1,0,1)	-	-	-	-	-
(1,1,0)	-	-	-	-	-
(1,1,1)	2.19	3	4	1.50	4

[18] See Table 4.2 and 4.1 for specific values, calculated by the *PyPhi* toolbox (Mayner et al. 2018).

Table 4.3: *Descriptive statistics of all state-dependent IIT-related values.*

I	μ	min	max	σ^2	γ_1
Φ^{Max}	2.0764	1.0000	3.2292	0.8802	-0.0855
$\Phi^{Max}_{Elements}$	2.3333	2.0000	3.0000	0.2667	-0.7071
$\Phi^{Max}_{Concepts}$	4.0000	2.0000	6.0000	2.8000	-0.2806
$\sum \varphi^{Max}$	1.6111	1.1667	2.1667	0.1741	0.1614
$\sum \varphi^{Max}_{Concepts}$	4.6667	4.0000	6.0000	0.6667	0.6261

The evidence of non-zero Φ^{Max} in a system allows for the analyzing of further metrics, for example, the average number of elements in the main complex is $\langle \Phi^{Max}_{Elements} \rangle = 2.67$. This number shows that there is at least one state where the main complex consists of less than three elements. If an elemental mechanism has no impact on other elements, it is not causal for the current experience and would also not be causal for the collective mental state. By this analysis, we could identify the behavior, which is not beneficial for the current state of the collective mind, e.g., reliability according to organizational mindfulness.

Another metric in IIT is the average amount of concepts in the main complex $\langle \Phi^{Max}_{Concepts} \rangle = 4.0$. This value represents the complexity of the social interrelating. Many concepts in a system can represent a tightly coupled system. Even if one concept drops out (e.g., a joint or individual mechanism), there are still concepts available that contain integrated conceptual information. Especially for organizational studies, it would be enlightening to analyze the concepts and their corresponding φ values. Concepts reveal the relationship between individuals and the current organizational mental state.

4.4 Discussion

It is difficult to imagine a collective mind and even more difficult to think of it as a real phenomenon. That is why it is more convenient to use the concept of a collective mind metaphorically when pointing to specific organizational behaviors. Mainly, it is the problem of a missing body (or the hidden physical structure) that prevents the observation at the organizational level, and studies have to rely on theoretical implications made by individual-level observations. Doing so creates a tension between the theoretical foundation of a collective mind in organizations and the methodological challenges and barriers to investigating them. However, we would miss out on a chance to better understand organizational cognition if these issues would not get attention in research. The integrated information theory brings us closer to understanding the collective mind and its physical representation. That enables to investigate the collective mind and how it is constructed by individual behavior, which is the basis for validating the quantity and quality of a collective mind and how it emerges from individual and joint behavior.

4.4.1 Contributions

The conceptual framework in this work can contribute to concepts in organizational cognition with social interrelating as a premise.[19] As there is no doubt that neurons play an essential role in our mental health, there is also no doubt that individuals play an essential role in organizational performance. However, only understanding the function of every single neuron would not enable us to understand the mind emerging from their activity. That is why this work proposes the examination of social interrelating, which leads to the emergence of organizational behavior. In particular, discussing the collective mind in terms of the integrated information theory can contribute to discussions about:

1. the collective mind and its relation to organizational performance,

2. the collective mind as an organizational boundary and,

[19] E.g., organizational mindfulness, organizational sense-making (Kudesia 2017, Stigliani & Ravasi 2012), collective interpretation (Gavetti & Warglien 2015), and collective identity (Lau & Rosenthal 2011).

3. the collective mind as an entity for memorizing organizational knowledge.

Besides the investigation of mental states at an organizational level, we can reduce the collective mind to its ability to represent supra-individual knowledge structures (Walsh 1995, Alavi & Leidner 2016). In that context, transactive memory systems (Wegner 1987, Wegner et al. 1985) or organizational memory (Walsh 1995) are closely related concepts. As Walsh & Ungson (1991) suggest, we should first investigate the structure of the organizational memory before identifying information processing and the function of the stored information for organizational performance. If memories were distributional, organizational memory would have several storage bins, e.g., individuals, culture, structures (Wegner 1987). This essay allows for the uncovering of the memory in one of those storage bins, the collective mind in the organization represented by integrated information.

Using IIT can help to find a bridge between different levels of analysis and how it relates to organizational performance. The above framework allows for the identification of the integrated information of the collective mind, but that alone can barely say anything about its function. If smart organizations would be reliable organizations, too (Weick & Roberts 1993), we can hardly tell if it is the collective mind that is causal to organizational performance (Morgeson & Hofmann 1999). Even in the human conscious mind, the footprints are neural activities, and it is not clear whether phenomenological experience has a function at all (Rosenthal 2008).

Besides the behavioral consequences of a collective mind, it is still relevant to finding a theory about the structure of supra-individual knowledge-structures to advance knowledge about micro-foundations (Barney & Felin 2013, Kozlowski et al. 2013, Kozlowski & Chao 2012). Hales & Chakravorty (2016) present a study to operationalize HRO practices in healthcare and show that mindful practices can improve organizational minding. Using mindfulness as a concept, we already accept that it can have positive effects on individual behavior (Kudesia 2019, Sutcliffe et al. 2016), which suggests at least a coexistence of organizational performance and subjective mental states. Still, it is hard to determine how individual behavior changes or is even changed by the collective mind. Especially multi-level research can help to advance knowledge in classic streams of organizational behavior (Porac & Tschang 2013, Theiner et al. 2010, Theiner & O'Connor 2010) and IIT can be applied in such studies.

As a consequence of measuring the collective mind by its integrated information, we can also determine its boundary, structure, and quality. According to IIT, interrelating social behavior with integrated information is a sufficient indicator of collective mental processes and not a metaphor at all. If the collective mind were more than a metaphor, then decisions made in such an organization cannot be reduced to the individual, too. Based on this idea, a number of discussions are already being held, such as group agency (List & Pettit 2006, List 2015), organizational decision-making (Aversa et al. 2018, Bavel et al. 2014), or the attention-based view of the firm (Stea et al. 2015). What unites these works is the discussion of the consequences of a collective mind, e.g., in legal or ethical issues. To contribute to those discussions, it would be possible to define the boundary of an organization based on organizational cognition (Santos & Eisenhardt 2005). Since this work describes how to determine whether an individual's behavior contributes to the collective mind, we can quantify its role in the organization, too, which offers a toolset to find the boundary of the organization.

4.4.2 Future research

What I describe above is a formal framework that makes it difficult to see how to investigate the collective mind in empirical research. Indeed, the framework contains a couple of pitfalls: It is challenging to decide which kind of behaviors to track, IIT's current algorithms assume fixed time-steps, and the discrete state of behavior can barely reflect its quality (how heedful the action is performed). I want to discuss those pitfalls briefly and offer ways to cope with them.

The variety of individuals and their behavior defines the set of mechanisms that would be the basis for the information-theoretic calculations. Even though the micro-foundations of the collective mind are interrelating behaviors, it is not clear which micro-foundations to observe in an organizational context. Hence, scientists would need to form hypotheses about sampling the micro-foundations of the collective mind carefully. Further, not only humans can be part of the organizational systems, but also machinery, especially regarding artificial intelligence (von Krogh 2018).

We know that neurons *fire* (change their state) in a specific rhythm and that their state is discrete (either they fire or not). Whereas, in an organization, some actions and behavioral patterns last longer than others, and its quality might fluctuate. For instance, Engel & Malone (2018) measure the collective intelligence and its relations to integrated information based on the quantity of communication during a task-solving experiment, but do not analyze the quality of the communicated information. However, measuring the quality of observed behavior can be substantial. From another perspective, determining the integrated information of particular behaviors can derive its quality for the collective mind without prejudice.

The conceptual framework in this work can be used to conduct multi-level studies investigating the dynamics between collective, joint, and individual mental states. However, it would require further methods to validate the role and power of a collective mind in organizations, which would be very difficult in ethnographies. In the first instance, laboratory experiments and computer simulations (Oliver et al. 2017, Davis et al. 2007) might be useful alternatives. A disadvantage for computer simulations is that either the behavior has to be programmed or the environment constraining the behavior that is crucial for collective behavior and organizational performance (Fischer et al. 2020). A blueprint on how to design such studies in organizational cognition would be the common-target game (Joyner 1970, Leavitt 1959), which was already used by Weick & Roberts (1993) to describe their ideas about an organizational mind.

4.4.3 Conclusion

Although the idea of a collective mind in organizations has existed for some time, we still know little about its specific structure and dynamics. I offer an approach to quantify the structure of a collective mind. The basis of the approach is a neuroscientific theory that explains how the conscious mind emerges from its micro-foundations. This framework strengthens the idea of the collective mind by validating that the theory offers more than an anthropomorphism. Still, investigating the phenomenon might require multi-level studies. Yet, even if the collective mind were only to have a tiny fraction of the complexity of a human mind, it can support scientists in organizational cognition in investigating supra-individual knowledge structures or organizational mental states.

4.5 References

Alavi, M. & Leidner, D. E. (2016), 'Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues', *MIS* **25**(1), 107–136.

Arico, A., Fiala, B., Goldbert, R. F. & Nichols, S. (2011), 'The Folk Psychology of Consciousness', *Mind & Language* **26**(3), 327–352.

Aversa, P., Cabantous, L. & Haefliger, S. (2018), 'When decision support systems fail: Insights for strategic information systems from Formula 1', *The Journal of Strategic Information Systems* **27**(3), 221–236.

Baars, B. J. (1988), 'A cognitive theory of consciousness', *NY: Cambridge University Press* .

Barney, J. & Felin, T. (2013), 'What are Microfoundations?', *Academy of Management Review* **27**(2), 138–155.

Bavel, J. J. V., Hackel, L. M. & Xiao, Y. J. (2014), The Group Mind: The Pervasive Influence of Social Identity on Cognition, *in* J. Decety & Y. Christen, eds, 'New Frontiers in Social Neuroscience', number November 2014 *in* 'Research and Perspectives in Neurosciences', Springer International Publishing, pp. 41–56.

Black, A. E. & McBride, B. B. (2013), 'Assessing high reliability practices in wildland fire management: An exploration and benchmarking of organizational culture', *USDA Forest Service - Research Note RMRS-RN* **RMRS-RN**(55), 1–17.

Block, N. (1995), The mind as the software of the brain, *in* S. S. D. N. Osherson, L. Gleitman, S. M. Kosslyn, S. Smith, ed., 'An Invitation to Cognitive Science', MIT Press, pp. 1–16.

Brown, K. W. & Cordon, S. (2009), Toward a Phenomenology of Mindfulness: Subjective Experience and Emotional Correlates, *in* F. Didonna, ed., 'Clinical Handbook of Mindfulness', number 1, Springer, pp. 59–81.

Bshary, R., Gingins, S. & Vail, A. L. (2014), 'Social cognition in fishes', *Trends in Cognitive Sciences* **18**(9), 465–471.

Carlo, J. L., Lyytinen, K. & Boland, R. J. (2012), 'Dialectics of Collective Minding: Contradictory Appropriations of Information Technology in a High-Risk Project', *MIS Quarterly* **36**(4), 1081–1108.

Chalmers, D. J. (1997), *The Conscious Mind*, Oxford University Press.

Clark, A. & Chalmers, D. J. (1998), The extended mind, *in* 'Analysis', Vol. 58, pp. 7–19.

Cornelissen, J. P. (2005), 'Beyond Compare: Metaphor in Organization Theory', *Academy of Management Review* **30**(4), 751–764.

Crick, F. & Koch, C. (2003), 'A framework for consciousness', *Nature Neuroscience* **6**(2), 119–126.
URL: *http://www.nature.com/articles/nn0203-119*

Davis, J. P., Eisenhardt, K. M. & Bingham, C. B. (2007), 'Developing Theory Through Simulation Methods', *Academy of Management Review* **32**(2), 480–499.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J. & Sergent, C. (2006), 'Conscious, preconscious, and subliminal processing: a testable taxonomy', *Trends in Cognitive Sciences* **10**(5), 204–211.

Dernbecher, S. & Beck, R. (2017), 'The concept of mindfulness in information systems research: A multi-dimensional analysis', *European Journal of Information Systems* **26**(2), 121–142.

Durkheim, E. (1895), *The Rules of Sociological Method*.

Edelman, G. (2006), 'Consciousness: The Remembered Present', *Annals of the New York Academy of Sciences* **929**(1), 111–122.

Engel, D. & Malone, T. W. (2018), 'Integrated information as a metric for group interaction', *PLOS ONE* **13**(10).

Fischer, D., Mostaghim, S. & Albantakis, L. (2020), 'How cognitive and environmental constraints influence the reliability of simulated animats in groups', *PLOS ONE* **15**(2), e0228879.

Fleck, L. (1983), *Genesis and Development of a Scientific Fact*, Vol. 11.

Fraher, A. L., Branicki, L. J. & Grint, K. (2017), 'Discovering How U.S. Navy SEALs Build Capacity for Mindfulness in High-Reliability Organizations (HROs)', *Academy of Management Discoveries* **3**(3), 239–261.

Freud, S. (1922), *Group Psychology and the Analysis of the Ego*.

Garnier, S., Gautrais, J. & Theraulaz, G. (2007), 'The biological principles of swarm intelligence', *Swarm Intelligence* **1**(1), 3–31.

Gavetti, G. & Warglien, M. (2015), 'A Model of Collective Interpretation', *Organization Science* **26**(5), 1263–1283.

Gebauer, A. (2017), *Kollektive Achtsamkeit organisieren*, Schäffer Poeschl.

Gordon, B. R. & Theiner, G. (2015), Scaffolded joint action as a micro-foundation of organizational learning, *in* C. B. Stone & L. M. Bietti, eds, 'Contextualizing Human Memory', number July, Psychology Press, pp. 154–186.

Halbwachs, M. (1952), *On Collective Memory*.

Hales, D. N. & Chakravorty, S. S. (2016), 'Creating high reliability organizations using mindfulness', *Journal of Business Research* **69**(8), 2873–2881.

Hutchins, E. (1989), 'Distributed Cognition', *IESBS Distributed Cognition* pp. 1–10.

Hutchins, E. (1995), 'How a Cockpit Remembers its Speed', *Cognitive Science* **19**, 265–288.

Jones, M. (1995), 'Organisational learning: Collective mind or cognitivist metaphor?', *Accounting, Management and Information Technologies* **5**(1), 61–77.

Joseph, J. & Gaba, V. (2019), 'Organizational structure, information processing, and decision making: A retrospective and roadmap for research', *Academy of Management Annals* pp. 1–83.

Joyner, R. C. (1970), *Computer simulation of individual concept learning in the three-person common target game*, Vol. 7.

Koch, C. (2018), 'What Is Consciousness?', *Nature* **557**(7704), S8–S12.

Koch, C., Massimini, M., Boly, M. & Tononi, G. (2016), 'Neural correlates of consciousness: progress and problems', *Nature Reviews Neuroscience* **17**(5), 307–321.

Kozlowski, S. W. J. & Chao, G. T. (2012), 'The Dynamics of Emergence: Cognition and Cohesion in Work Teams', *Managerial and Decision Economics* **33**(5-6), 335–354.

Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T. & Kuljanin, G. (2013), 'Advancing Multilevel Research Design: Capturing the Dynamics of Emergence', *Organizational Research Methods* **16**(4), 581–615.

Kudesia, R. S. (2017), Organizational Sensemaking, *in* 'Oxford Research Encyclopedia of Psychology', Vol. 53, Oxford University Press, pp. 286–305.

Kudesia, R. S. (2019), 'Mindfulness as metacognitive practice', *Academy of Management Review* **44**(2), 405–423. **URL:** *http://journals.aom.org/doi/10.5465/amr.2015.0333*

Lau, H. & Rosenthal, D. (2011), 'Empirical support for higher-order theories of conscious awareness', *Trends in Cognitive Sciences* **15**(8), 365–373.

Le Bon, G. (1896), *The Crowd: A Study of the Popular Mind*.

Leavitt, H. J. (1959), 'Task ordering and organizational development in the common target game', *Behavioral Science* **5**(3), 233–239.

List, C. (2015), What is it like to be a group agent?

List, C. & Pettit, P. (2006), 'Group Agency and Supervenience', *The Southern Journal of Philosophy* **44**(May 2005), 1–22.

Long, D. M. (2005), *Mind – Introduction to Cognitive Science*, Vol. 15.

March, J. G. (1962), 'The Business Firm as a Political Coalition', *The Journal of Politics* **24**(4), 662–678.

March, J. G. (1996), 'Continuity and change in theories of organizational action', *Administrative Science Quarterly* **41**(2), 278–287.

March, J. G. & Simon, H. A. (1958), *Organizations*, Wiley, New York, NY.

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R. & Tononi, G. (2018), 'PyPhi: A toolbox for integrated information theory', *PLOS Computational Biology* **14**(7), e1006343.

Morgeson, F. P. & Hofmann, D. A. (1999), 'The Structure and Function of Collective Constructs: Implications for Multilevel Research and Theory Development', *Academy of Management Review* **24**(2), 249–265.

Nagel, T. (1974), 'What Is It Like to Be a Bat?', *The Philosophical Review* **83**(4), 435.

Nelson, T. O. (2002), 'Consciousness, Self-Consciousness, and Metacognition', *Consciousness and Cognition* **9**(2), 220–223.

Nguyen, T. T. & Belk, R. W. (2007), 'This We Remember: Consuming Representation via the Web Posting of War Photographs', *Consumption Markets & Culture* **10**(3), 251–291.

Oizumi, M., Albantakis, L. & Tononi, G. (2014), 'From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* **10**(5), 1–25.

Oliver, N., Senturk, M., Calvard, T. S., Potocnik, K. & Tomasella, M. (2017), 'Collective Mindfulness, Resilience and Team Performance', *Academy of Management Proceedings* **2017**(1).

Orton, J. D. & Weick, K. E. (1990), 'Loosely Coupled Systems: A Reconceptualization', *Academy of Management Review* **15**(2), 203–223.

Porac, J. & Tschang, F. T. (2013), 'Unbounding the Managerial Mind: It's Time to Abandon the Image of Managers As 'Small Brains'', *Journal of Management Inquiry* **22**(2), 250–254.

Rerup, C. (2004), 'Variations in Organizational Mindfulness', *Academy of Management Proceedings* **2004**(1), B1–B5.

Rerup, C. & Levinthal, D. A. (2014), Situating the Concept of Organizational Mindfulness: The Multiple Dimensions of Organizational Learning, *in* G. Becke, ed., 'Mindful Change in Times of Permanent Reorganization', number April 2016 *in* 'CSR, Sustainability, Ethics & Governance', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 131–145.

Rosenthal, D. M. (2005), *Consciousness and the Mind*, Oxford University Press UK.

Rosenthal, D. M. (2008), 'Consciousness and its function', *Neuropsychologia* **46**(3), 829–840.

Sandelands, L. E. & Stablein, R. E. (1987), 'The Concept of Organization Mind', *Research in the Sociology of Organizations* **5**, 135–161.

Santos, F. M. & Eisenhardt, K. M. (2005), 'Organizational Boundaries and Theories of Organization', *Organization Science* **16**(5), 491–508.

Schmitt, R. (1959), 'Husserl's Transcendental-Phenomenological Reduction', *Philosophy and Phenomenological Research* **20**(2), 238.

Searle, J. R. (1980), 'Minds, brains, and programs', *Behavioral and Brain Sciences* **3**, 417–424.

Searle, J. R. (1991), 'Consciousness , Unconsciousness Intentionality', *Philosophical Issues* **1**, 45–66.

Searle, J. R. (2000), 'Consciousness', *Intellectica* **31**, 85–110.

Searle, J. R. (2005), 'What is an institution?', *Journal of Institutional Economics* **1**(1), 1–22.

Shannon, C. E. (1948), 'A Mathematical Theory of Communication', *Bell System Technical Journal* **27**(3), 379–423.

Shapira, Z. (2011), '"I've Got a Theory Paper—Do You?": Conceptual, Empirical, and Theoretical Contributions to Knowledge in the Organizational Sciences', *Organization Science* **22**(5), 1312–1321.

Simon, H. A. (1947), *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organizations*, New York: Macmillan.

Stea, D., Linder, S. & Foss, N. J. (2015), Understanding Organizational Advantage: How the Theory of Mind Adds to the Attention-Based View of the Firm, pp. 277–298.

Stigliani, I. & Ravasi, D. (2012), 'Organizing thoughts and connecting brains: Material practices and the transition from individual to group-level prospective sensemaking', *Academy of Management Journal* **55**(5), 1232–1259.

Sutcliffe, K. M., Vogus, T. J. & Dane, E. (2016), 'Mindfulness in Organizations: A Cross-Level Review', *Annual Review of Organizational Psychology and Organizational Behavior* **3**(1), 55–81.

Tegmark, M. (2015), 'Consciousness as a state of matter', *Chaos, Solitons & Fractals* **76**, 238–270.

Theiner, G., Allen, C. & Goldstone, R. L. (2010), 'Recognizing group cognition', *Cognitive Systems Research* **11**(4), 378–395.

Theiner, G. & O'Connor, T. (2010), The Emergence of Group Cognition, *in* 'Emergence in science and philosophy', pp. 92–132.

Tomasello, M. (2014), *A natural history of human thinking*, Harvard University Press.

Tononi, G. (2004), 'An information integration theory of consciousness', *BMC Neuroscience* **5**(1), 42.

Tononi, G. (2012), 'Integrated information theory of consciousness: an updated account', *Archives Italiennes de Biologie* **150**, 290–326.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016), 'Integrated information theory: from consciousness to its physical substrate.', *Nature reviews. Neuroscience* **17**(7), 450–61.

Tononi, G. & Koch, C. (2008), 'The neural correlates of consciousness: An update', *Annals of the New York Academy of Sciences* **1124**, 239–261.

Vendelo, M. T. & Rerup, C. (2019), 'Collective Mindfulness in a Regenerating Organization: Ethnographic Evidence from Roskilde Festival', *Safety Science* .

Voegelin, E. (2000), *Order and History, Plato and Aristotle: Volume III*, University of Missouri Press, Columbia and London.

Vogus, T. J. & Sutcliffe, K. M. (2007), 'The Safety Organizing Scale', *Medical Care* **45**(1), 46–54.

von Krogh, G. (2018), 'Artificial Intelligence in Organizations: New Opportunities for Phenomenon-Based Theorizing', *Academy of Management Discoveries* **4**(4), 404–409.
URL: *http://journals.aom.org/doi/10.5465/amd.2018.0084*

Walsh, J. & Louvre, A. (1988), Introduction, *in* 'Tell Me Lies About Vietnam: Cultural Battles for the Meaning of the War', Open University Press, Philadelphia, pp. 1–29.

Walsh, J. P. (1995), 'Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane', *Organization Science* **6**(3), 280–321.

Walsh, J. & Ungson, G. R. (1991), 'Organizational Memory', *Academy of Management Review* **16**(1), 57–91.

Wegner, D. M. (1987), Transactive memory: A contemporary analysis of the group mind, *in* 'Theories of group behavior', Springer, New York, chapter 9, pp. 185–208.

Wegner, D. M., Giuliano, T. & Hertel, P. T. (1985), Cognitive Interdependence in Close Relationships, *in* 'Compatible and Incompatible Relationships', Springer New York, New York, NY, chapter 11, pp. 253–276.

Weick, K. E. (1976), 'Educational Organizations as Loosely Coupled Systems', *Administrative Science Quarterly* **21**(1), 1.

Weick, K. E. (1989), 'Theory Construction as Disciplined Imagination', *Academy of Management Review* **14**(4), 516–531.

Weick, K. E. & Roberts, K. H. (1993), 'Collective Mind in Organizations: Heedful Interrelating on Flight Decks', *Administrative Science Quarterly* **38**(3), 357.

Weick, K. E. & Sutcliffe, K. M. (2015), *Managing the unexpected*, Vol. 66, 3rd edn, John Wiley & Sons, Inc., New Jersey.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2008), 'Organizing for High Reliability: Process of Collective Mindfulness', *Crisis Management* **3**, 31–66.

Will, T. E. (2016), 'Flock Leadership: Understanding and influencing emergent collective behavior', *Leadership Quarterly* **27**(2), 261–279.

Chapter 5

Wirtschaftsprüfung im Zeitalter der Digitalisierung

Summary

Auditors are not known to take in a leadership role when it comes to innovations since they orient their operation on the current law and obligations. However, digital transformation can disrupt this traditional sector. Their clients and even law can change so that the auditors are required to implement a digitalization strategy to validate them. Even if this is unlikely in the short term, auditors can use new technology to optimize and supplement their day-to-day work. This essay analyzes the current audit process compared to digitalization and determines the potential of digitalization in auditing. Further, it discusses contradictions between the auditor's independence and the possibility of services that lay beyond the actual audit contract. We find that auditors can profit from the digitalization, assuming they design their strategy proactive. If they behave as usual – to react on market or law changes – they might risk losing their clients to a competitor, since the speed of upcoming innovations is increasing, while the pressure on prices is high. Generally, we believe that the audit profession will not die out but will undergo a significant change.

Keywords: Digital transformation, industry 4.0, auditing, financial accounting

Co-Author:	Benedikt Downar (see the Appendix for the declaration of the individual contribution).
Current-Status:	*Published*, see: Downar B, Fischer D. Wirtschaftsprüfung im Zeitalter der Digitalisierung. In: Obermeier R, editor. Handbuch Industrie 4.0 und Digitale Transformation. 1st ed. Gabler Verlag; 2019. doi:10.1007/978-3-658-24576-4
Permission:	Reprinted by permission from: Springer Nature, 2019. License Number: 4841771162930.

5.1 Einleitung

Historisch betrachtet ist die Wirtschaftsprüfung nicht als Vorreiter für Innovationen bekannt (Alles 2015, Dai & Vasarhelyi 2016). Viel eher orientiert sich die Arbeit der Wirtschaftsprüfer[1] an den bestehenden gesetzlichen und berufsständischen Vorschriften. Änderungen in der Arbeit der Wirtschaftsprüfer ergeben sich daher eher als Folge geänderter rechtlicher Anforderungen und weniger als Reaktion auf neue technische Möglichkeiten.

Zu den zentralen Aufgaben der Wirtschaftsprüfer gehört die Durchführung von betriebswirtschaftliche Prüfungen, wie die Jahresabschlussprüfung (§ 2 Abs. 1 Wirtschaftsprüferordnung (WPO)). Im Zuge der vierten industriellen Revolution (Obermaier 2017) könnte sich diese Tätigkeit deutlich verändern. Als Folge der Digitalisierung und der Vernetzung von Prozessen werden Wirtschaftsprüfer zunehmend mit Themen wie Artificial Intelligence, Big Data, Cloud Computing und Audit Data Analytics konfrontiert. Zudem werden traditionelle Prüfungshandlungen, unter anderem die stichprobenartige Auswertung von Belegen, durch die Menge an zu prüfenden Sachverhalten aufwendiger und verlieren an Aussagekraft (Ruhnke 2017).

Der vorliegende Beitrag beschäftigt sich daher mit den Chancen und Risiken der Digitalisierung für die Wirtschaftsprüfung. Da bisher nur wenige praktische Anwendungsfälle existieren[2], erfolgt insbesondere eine Analyse auf Basis der theoretischen Anwendungsmöglichkeiten. Der Beitrag befasst sich dabei sowohl mit dem eigentlichen Prüfungsprozess, den Möglichkeiten in anderen Tätigkeitsbereichen wie der Beratung, der Praxisorganisation, als auch mit der Bedeutung für die Berufsausbildung und möglichen Effekten für die Marktstruktur. Ebenso wird diskutiert wieso die Wirtschaftsprüfung besonders gefordert ist Innovationen mit disruptivem Potenzial (Obermaier 2017) zu entwickeln und wie der Wandel beschleunigt werden könnte.

5.2 Der Prüfungsprozess im Spannungsfeld der Digitalisierung

5.2.1 Objekt, Ansatz und Durchführung der Abschlussprüfung

Gemäß § 316 Abs. 1 Satz 1 Handelsgesetzbuch (HGB) sind der Jahresabschluss und Lagebericht von Kapitalgesellschaften, die mindestens zwei von drei Kriterien nach § 267 Abs. 1 HGB (Bilanzsumme, Arbeitnehmerzahl und Umsatzerlöse) in zwei aufeinanderfolgenden Geschäftsjahren erfüllen, durch einen Abschlussprüfer zu prüfen.[3] Damit soll die Verlässlichkeit und Glaubhaftigkeit der bereitgestellten Informationen bestätigt beziehungsweise erhöht werden (Institut der Wirtschaftsprüfer (IDW) Prüfungsstandard (PS) 200 Tz. 8). Die Abschlussprüfung trägt somit wesentlich zum Investorenschutz bei. Die Prüfungshandlungen sind dabei so auszulegen, dass Unrichtigkeiten und Verstöße mit einem wesentlichen Einfluss auf die Vermögens-, Finanz-, und Ertragslage aufgedeckt werden (§ 317 Abs. 1 HGB). Aus dieser Zielsetzung ergibt sich, dass die Abschlussprüfung keine Vollprüfung ist, sondern dem Grundsatz der Wirtschaftlichkeit folgt (IDW PS 200 Tz. 19 und 21). Die Prüfungstätigkeit ist so zu bemessen, dass mit hinreichender Sicherheit die Ordnungsmäßigkeit der Rechnungslegung beurteilt werden kann (Ewert & Wagenhofer 2015). Die Fokussierung auf wesentliche Fehler erfolgt dabei nach dem Konzept des risikoorientierten Prüfungsansatzes (Quick 1996, Brösel et al. 2015). Das Prüfungsrisiko ergibt sich dabei wie folgt:

```
Prüfungsrisiko = Inhärentes Risiko × Kontrollrisiko × Entdeckungsrisiko
```

Inhärentes Risiko beschreibt die Fehlerwahrscheinlichkeit die sich aus der Branche des zu prüfenden Unternehmens, den damit assoziierten Risiken sowie der Wahrscheinlichkeit einer bewussten Beeinflussung durch das Unternehmen ergibt. Kontrollrisiko bezeichnet das Risiko, dass wesentliche Fehler nicht durch interne Kontrollsysteme entdeckt bzw. verhindert werden. Das Entdeckungsrisiko bezeichnet das Risiko, dass wesentliche Fehler

[1] Zur Vereinfachung wird im Folgenden nicht zwischen Wirtschaftsprüfern und Wirtschaftsprüferinnen differenziert.

[2] Einen bekannten Ansatz verfolgt PricewaterhouseCoopers mit ihrem Produkt Halo (Bartmann et al. 2018).

[3] Regelungen für bestimmte Personenhandelsgesellschaften nach § 3 Publizitätsgesetz (PublG) sowie die Prüfungspflicht für Konzernabschlüsse werden an dieser Stelle nicht vertieft. Die dargestellten Themen gelten für andere Abschlussprüfungen äquivalent.

durch die durchgeführten Prüfungshandlungen nicht aufgedeckt werden. Während inhärentes Risiko und Kontrollrisiko nicht direkt vom Wirtschaftsprüfer beeinflusst werden können, wird das Entdeckungsrisiko durch den Umfang der durchgeführten Prüfungshandlungen bestimmt (Ewert & Wagenhofer 2015).

Prüfungshandlungen sind dabei so zu gestalten, dass ausreichende und angemessene Prüfungsnachweise erlangt werden (IDW PS 300 Tz. 6). Diese Anforderungen beziehen sich dabei auf die Quantität und Qualität der Nachweise. Hinsichtlich der Beurteilung der Qualität ist die inhaltliche Sachdienlichkeit sowie die Verlässlichkeit eines Nachweises zu beachten. Die Verlässlichkeit ergibt sich unter anderem aus der Art (z. B. Original versus Kopie), der Herkunft (interner versus externer Nachweis) und dem Bezug (z. B. durch Wirtschaftsprüfer oder Unternehmen beigebracht) der Prüfungsnachweise (IDW PS 300 Tz. 39). Nachweise sind dabei nicht auf interne Informationen des zu prüfenden Unternehmens beschränkt, auch Informationen von externen Quellen, wie Banken oder Lieferanten, können und sollen als Prüfungsnachweise berücksichtigt werden (IDW PS 302 n.F.).

Wie die einzelnen Prüfungsrisiken beurteilt beziehungsweise gewichtet werden, liegt letztendlich im Ermessen des verantwortlichen Wirtschaftsprüfers. Aus Sicht der theoretischen Informatik lässt sich dieses Vorgehen auf ein meist regelbasiertes Vorgehen reduzieren. Einer Prüfungshandlung liegt dabei eine vordefinierte Regel zu Grunde, welche beispielsweise aus einem Prüfungsstandard abgeleitet wird. Sollte eine Auffälligkeit auftreten, kann so eine kausale Kette zu geltenden Vorschriften hergestellt werden und eine angemessene Reaktion daraus abgeleitet werden (Haas 2017). Dieses Vorgehen ist bereits heute in Form von digitalen Expertensystemen im Einsatz (Grosan & Abraham 2011). Bei der Risikoeinschätzung ist allerdings zu beachten, dass digitale Regelsysteme fundamental anders aufgebaut sind als menschliche Regelsysteme (Pomerol 1997). Entscheidungsmodelle im Bereich Deep Learning und Big Data entwickeln beispielsweise selbständig Regelsysteme aus einer vordefinierten Lernmenge oder einer mathematischen Fitnessfunktion[4]. Beispielsweise liegt Bilderkennungssoftware eine Datenbank zugrunde, wobei zunächst jedes Bild von einem Menschen beschrieben wurde (z. B. Zuordnung von Namen der abgebildeten Personen). Auch wenn die Software dann ein neu aufgenommenes Bild richtig erkennen kann, d.h. bereits bekannte Personen richtig identifiziert, handelt es ich dabei lediglich um ein mathematisches Zuordnungsverfahren. Um den risikoorientierten Prüfungsansatz in ein digitales System zu übertragen, müssten nicht nur die vorhandenen Prüfungsstandards und Gesetzte in einen regelbasierten Algorithmus oder eine Datenbank transferiert werden, sondern auch der Handlungsspielraum und der Erfahrungsschatz der individuellen Wirtschaftsprüfer.

5.2.2 Digitalisierung in der Abschlussprüfung heute

Die Notwendigkeit einer Weiterentwicklung der Abschlussprüfung ergibt sich aus der zunehmenden Digitalisierung der mandantenspezifischen Rechnungslegungssysteme und der Menge und Komplexität der zu prüfenden Sachverhalte. Beispielsweise nutzen Unternehmen immer häufiger standardisierte ERP Systeme (Goldshteyn et al. 2013). Diese Systeme erlauben einen strukturierten Datenabzug der Geschäftsvorfälle. Hieraus resultieren zahlreiche neue Möglichkeiten zur Analyse mandantenspezifischer Daten welche über den aktuellen Stand der IT-gestützten Prüfungstechniken hinausgehen (Ruhnke 2017, Goldshteyn et al. 2013).

Bereits in den letzten Jahren wurden vom IDW verschiedene Normen erlassen, die Wirtschaftsprüfer im Umgang mit IT-Systemen unterstützen sollen und einen Rahmen für den Einsatz von Softwareprogrammen im Rahmen der Abschlussprüfung abstecken. Hierbei sind besonders IDW PS 330, IDW Stellungnahme zur Rechnungslegung (RS) Fachausschuss für Informationstechnologie (FAIT) 1 und IDW Prüfungshinweis (PH) 9.330.3 zu nennen. Diese Normen regeln die Pflichten im Rahmen der Prüfung von IT-Systemen sowie die Möglichkeiten zum Einsatz digitaler Datenanalysen zur Erlangung von Prüfungsnachweisen. Zu beachten ist aber, dass die Normen ihren Ursprung in den Jahren 2002 (IDW PS 330; IDW RS FAIT 1) beziehungsweise 2010 (IDW PH 9.330.3) haben und folglich im Kontext ihrer Zeit zu sehen sind.

Sofern die Buchführung unter Verwendung von IT-gestützten Rechnungslegungssystemen erfolgt, ist auch das hierzu verwendete System Gegenstand der Abschlussprüfung. Zielsetzung der IT-Systemprüfung nach IDW PS 330 ist die Ordnungsmäßigkeit und die Sicherheit dieses Systems zu überprüfen und wesentliche inhärente IT-

[4] Eine Fitnessfunktion ist die Zielfunktion eines selbstlernenden Systems, das sich iterativ einem Funktionsoptimum annähert.

Fehlerrisiken zu identifizieren. Ordnungsmäßigkeit ist dabei, bezugnehmend auf die Grundsätze ordnungsgemäßer Buchführung (GoB), anhand von sechs Kriterien definiert (IDW RS FAIT 1 Tz. 25-32):

- Vollständigkeit: die lückenlose Erfassung aller Geschäftsvorfälle und Vermeidung von Mehrfachbuchungen.
- Richtigkeit: die Sachverhalte müssen inhaltlich richtig und in Übereinstimmung mit den rechtlichen Vorgaben abgebildet werden.
- Zeitgerechtheit: die Sachverhalte sind unmittelbar nach Entstehung abzubilden und den korrekten Perioden zuzuordnen.
- Ordnung: die Darstellung der Geschäftsvorfälle in sachlicher (Kontenfunktion) und zeitlicher (Journalfunktion) Ordnung muss gewährleistet sein.
- Nachvollziehbarkeit: ein sachverständiger Dritter muss sich in angemessener Zeit einen Überblick über die Geschäftsvorfälle machen und die angewandten Buchführungs- und Rechnungslegungsverfahren nachvollziehen können.
- Unveränderlichkeit: es dürfen keine Änderungen durchgeführt werden, die den ursprünglichen Inhalt nicht mehr feststellbar machen. Änderungen an den Generierungs- und Steuerungsdaten sind zu dokumentieren.

Diese Kriterien richten sich an Unternehmen, die klassische ERP-Systeme, wie zum Beispiel von SAP oder Oracle, nutzen. Hierbei ist allerdings zu beachten, dass diese Systeme in erster Linie von Menschen bedient und kontrolliert werden. Mittelfristig wäre es denkbar, Abteilungen komplett zu digitalisieren. Beispielsweise könnte in produzierenden Unternehmen der komplette Einkaufsprozess vollständig automatisiert ablaufen: Intelligente Systeme scannen dabei den aktuellen Rohstoffmarkt, ermitteln die zukünftige Absatzprognose und kaufen – autonom – zum optimalen Preis ein (Uygun & Ilie 2018). Aus Sicht der Wirtschaftsprüfer birgt dies neue Risiken für die Abschlussprüfung. Neben dem Risiko, dass Prozesse zur Übertragung von Sachverhalten gestört und verändert werden könnten, wäre es auch denkbar, dass Dritte künstlichen Intelligenzen zu ihren Gunsten manipulieren. Schon jetzt generieren Bots zwischen 6 und 7 Milliarden US-Dollar Schaden pro Jahr durch das Aktivieren von digitalen Werbeanzeigen (WhiteOps & ANA 2017). Auch die bereits erwähnten Bilderkennungsdienste können so beeinflusst werden, dass kein sinnvolles Ergebnis mehr zu erwarten ist (Nguyen et al. 2015). Entsprechend sollte hinterfragt werden, ob die zuvor genannten Kriterien ausreichend sind oder nicht dahingehend erweitert werden müssten, dass die digitalen Prozesse, welche die Geschäftsvorfälle produzieren, selbst geprüft werden[5] (Dai & Vasarhelyi 2016).

Die Anforderung der Sicherheit gilt dann als gegeben, wenn ein Sicherheitskonzept implementiert wurde, dass die folgenden Anforderungen erfüllt (IDW RS FAIT 1 Tz. 23 ff.):

- Vertraulichkeit: von Dritten erlangte Daten werden nicht unberechtigt veröffentlicht oder weitergegeben.
- Integrität: Daten, Infrastruktur und Anwendung stehen vollständig und richtig zur Verfügung und sind vor Manipulation geschützt.
- Verfügbarkeit: die ständige Verfügbarkeit der Infrastruktur, Anwendungen und Daten müssen gewährleistet sein. Für Ausfälle müssen Back-up-Verfahren bereitstehen und die Lesbarkeit der Buchführung möglich sein.
- Autorisierung: durch die Einrichtung von Zugriffsschutzmaßnahmen muss sichergestellt sein, dass nur autorisierte Personen Zugriff auf das System haben und auch nur Zugriffsrechte in klar definierten Grenzen.
- Authentizität: d.h. Geschäftsvorfälle müssen eindeutig einem Verursacher zugeordnet werden können.
- Verbindlichkeit: Transaktionen dürfen nicht abstreitbar sein und führen Rechtsfolgen bindend herbei.

Üblicherweise dokumentieren Unternehmen Geschäftsvorfälle durch digitale Belege und Papierbelege. In solch einer Situation hat ein Wirtschaftsprüfer wenige Probleme oben genannte Kriterien zu überprüfen. Immer häufiger sind allerdings Unternehmen, die einen Teil der Dokumente rein digital ablegen. Hier ist es schwieriger zu überprüfen, ob die Belege integer oder vertraulich sind. Spuren lassen sich in digitalen Datenbanken leichter verwischen

[5] Darüber hinaus ist zu erwarten, dass durch einen automatisierten Einkauf mehr Transaktionen in kürzerer Zeit getätigt werden können. Dies wird sich ebenfalls auf die Abschlussprüfung auswirken.

als auf gedruckten Dokumenten. Daher muss sichergestellt sein, dass diese Datenbanken geltenden Sicherheitsstandards[6] entsprechen (IDW PS 330). Darüber hinaus können bei der Digitalisierung von Papierbelegen Fehler entstehen, z. B. durch fehlerhafte Scanprogramme (Gerber 2013).

Mit der Einführung von mehr und mehr vernetzten Systemen ist es allerdings insgesamt fraglich, ob die aktuelle IT-Systemprüfung überhaupt ausreichend ist. Zum Beispiel könnten Daten nicht mehr dediziert im Serverraum der Unternehmen liegen, sondern auf Cloud-Speicher von Drittanbietern verteilt werden (Dai & Vasarhelyi 2016). Es ist somit zu prüfen, inwieweit IT-Dienstleister die oben genannten Kriterien erfüllen können. Als Alternative zu den bisherigen Datenbanken könnte beispielsweise die Blockchain-Technologie zum Einsatz kommen (Richins et al. 2017). Alle Arten von Transaktionen werden dabei in einer zusammenhängenden und untereinander verknüpften Kette abgebildet, sodass Beeinflussungen nahezu ausgeschlossen sind. Auf diese Weise könnte das Prüfungsrisiko im Bereich der IT-Systeme reduziert werden.

Neben der Prüfung mandantenspezifischer IT-Systeme kommen auch heute schon verschiedene IT-gestützte Prüfungstechniken auf Seiten der Wirtschaftsprüfer zum Einsatz. Die Bereitstellung solcher Tools erfolgt dabei häufig durch standardisierte Prüfsoftware, wie z. B. ACL Analytics und Audicon IDEA[7]. Der Einsatz solcher Computer-Assisted Audit Tools (CAAT) ist auf allen Stufen des Prüfungsprozesses üblich.[8] Hierbei ist allerdings zu beachten, dass IT-gestützte Datenanalysen, nach IDW PH 9.330.3 Tz. 5 und 18, allein nicht ausreichend sind um ein hinreichendes Maß an Prüfungssicherheit zu erlangen. Die IT-gestützte Datenanalyse wird auch bereits aktuell als wichtig und nützlich angesehen. Der Großteil der Nutzer setzt die dedizierten CAATs aber primär für die Haupt- und Nebenbuchprüfung ein (PwC 2017), also zur Automatisierung und Ergänzung originärer Prüfungstätigkeiten. Die Begründung für die Beschränkung auf klassische Einsatzfelder liegt vermutlich darin, dass die in IDW PS 330 und IDW PH 9.330.3 exemplarisch angeführten Anwendungsfälle eher der Arbeitsorganisation (z. B. Projektplanungsanwendung, Präsentationsprogramme, Tabellenkalkulation), der Vorbereitung weiterer Prüfungshandlungen (z. B. die maschinelle Berechnung von Stichproben) sowie den grundlegenden statistischen Auswertungen (Zeitreihen-, Trend-, Abweichungs- und Strukturanalysen sowie Berechnungen, Auswertungen und Aufbereitungen) zuzuordnen sind. Eine Analyse von elf gängigen Softwarepaketen durch die Unternehmensberatung Roger Odenthal und Partner Odenthal (2017) zeigt zudem, dass lediglich zwei der untersuchten Softwarepakete fortgeschrittene statistische Verfahren wie Korrelations- und Regressionsanalysen unterstützen. Darüberhinausgehende Verfahren, die dem Prüfer evidenzbasierte Entscheidungen auf Grundlage von intelligenter Datenanalyse liefern, gehören nicht zum Standardumfang. Insgesamt lässt sich also festhalten, dass sich der Einsatz von Prüfsoftware eher am Grundsatz der Wirtschaftlichkeit orientiert (Effizienzsteigerung) und weniger eine Ausreizung der technischen Potenziale angestrebt wird (Effektivitätssteigerung).

5.2.3 Potenziale für den Prüfungsprozess

Ein gern verwendeter Begriff im Silicon Valley ist die Disruption. Eine Technologie mit disruptiven Potenzial beschreibt eine Innovation, welche eine bestehende Technologie, bestehende Prozesse oder auch vollständige Dienstleistungen komplett verdrängt (Bower & Christensen 1995). Wie eine solche Disruption einen Weltkonzern auslöschen kann, zeigt das Beispiel der Eastman Kodak Company (kurz: Kodak). Traditionelle Produkte wurden hier der Digitalfotografie vorgezogen und das Unternehmen schaufelte sich so letztendlich sein eigenes Grab (Munir 2012).

Die im vorherigen Punkt geschilderten Normen zur IT-Unterstützung der Prüfungshandlung zielen lediglich auf einen effizienteren Ablauf von traditionellen Prozessen ab. Es ist entsprechend zu hinterfragen, ob noch Wirtschaftsprüfer für den Prüfungsprozess gebraucht werden, wenn Unternehmen mehr und mehr intelligente(re) ERP-Systeme einsetzen. Daher ist naheliegend zu erwarten, dass der Berufsstand eigene Innovationen vorantreibt um sich an die geänderten technischen Möglichkeiten anzupassen.

---

6 Z. B. ISO/IEC 27001.

7 Für weitergehende Erläuterungen vgl. `https://www.acl.com/` beziehungsweise `https://audicon.net/`.

8 Die Regelungen des Bundesdatenschutzgesetzes, insb. § 26 BDSG auf Grund von § 320 Abs. 1 Satz 2 und Abs. 2 HGB, finden bei der Verwendung von IT-gestützten Prüfungstechniken keine Anwendung.

Ein Hemmnis hierbei könnten, abgeleitet von Curtis & Schulman (2006), allerdings die teilweise restriktiven gesetzlichen Datenschutz- und Sicherheitsrichtlinien darstellen. Insbesondere resultiert aus der Pflicht zur Verschwiegenheit, dass Mandantendaten nur fallbezogen verwendet werden dürfen und somit nicht für die Entwicklung von neuer Verfahren dienen können. Wie Curtis & Schulman (2006) am Beispiel des Innovator‘s Dilemma nach Christensen (1997) beschreiben, stellen erhöhte regulatorische Anforderungen eine Barriere für Innovationen mit disruptiven Potenzial dar (siehe Abbildung 5.1).[9] Hierbei wird grundlegend zwischen Innovationen an etablierten Produkten (sustaining innovation) und Innovationen mit disruptiven Potential unterschieden. Sustaining innovations zielen darauf ab bestehende Kundenwünsche, durch Modifikation bestehender Produkte, weiterhin zu erfüllen. Innovationen mit disruptiven Potenzial zielen hingegen auf vollständig neue Produkte und Funktionen ab, die anfangs oft nur eine (exklusive) Minderheit bedienen. Dies liegt darin begründet, dass der Mehrwert dieser Innovation oft nicht sofort ersichtlich ist. Der Beförderungsdienst Uber und das Apple iPhone sind bekannte Fallbeispiele hierfür. Uber verdrängt in den Vereinigten Staaten den traditionellen Taximarkt und Apple hat mit dem iPhone den Mobilfunkmarkt revolutioniert.[10]

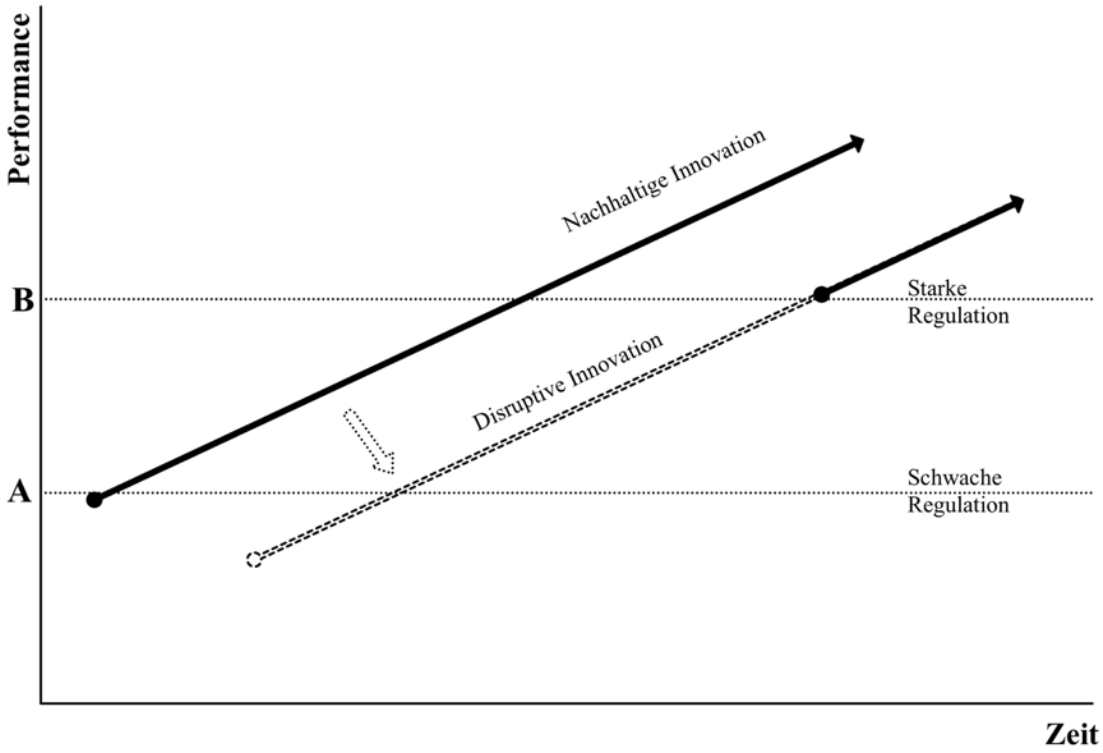


Abbildung 5.1: *Modifizierte Version des Innovator's Dilemma von Christensen (1997) nach Curtis & Schulman (2006).*

Curtis & Schulman (2006) betonen, dass in stark regulierten Branchen ein höherer Aufwand nötig ist, um die Marktreife und eine Produktzulassung zu erreichen. Außerdem ist es in diesen Branchen unwahrscheinlicher, dass Kunden Produkte nachfragen, die wesentlich mehr leisten als gesetzliche Vorschriften verlangen.

Überträgt man das Beispiel von Curtis & Schulman (2006) auf die Wirtschaftsprüfung, so lässt sich der Innovationsmangel dadurch erklären, dass die hohen Anforderungen für die Abschlussprüfung die Etablierung neuer Technologien grundlegend erschwert. Darüber hinaus bestehen möglicherweise wenige Anreize für hohe Investitionsausgaben, da bereits heute der Wettbewerb um Mandate sehr hoch ist und entsprechend eine Finanzierung der Ausgaben über höhere Prüfungshonorare nur begrenzt möglich ist (Lückmann 2014). Mögliche neue Startups im Bereich der Wirtschaftsprüfung, welche oft Vorreiter für disruptive Innovationen sind (Christensen 1997),

[9] Curtis & Schulman (2006) nutzen als Beispiel für stark regulierte Branchen das Gesundheitswesen.
[10] Vgl. diesbezüglich auch Obermaier (2017).

werden durch diese Markteintrittsbarrieren zusätzlich ausgebremst. Von technologischer Seite betrachtet sind die Potenziale für den Prüfungsprozess und somit auch für die zu prüfenden Unternehmen aber vielfältig und könnten sich auf alle Stufen des Prüfungsprozesses und auch weit über den klassischen Prüfungsprozess hinaus erstrecken. Im Folgenden werden daher verschiedene mögliche Innovationen für die einzelnen Stufen des Prüfungsprozesses dargestellt.[11]

5.2.4 Auftragsannahme und Planung

Im Rahmen der Auftragsannahme ist es für Wirtschaftsprüfer essentiell mögliche Risiken, die mit einem neuen Mandat verbunden sein können, zu identifizieren. Um dieses Verständnis zu erlangen, kann eine Auswertung nicht nur interner, sondern auch externer Daten von großem Nutzen sein (Issa et al. 2016, No & Vasarhelyi 2017). Künstliche Intelligenzen[12] können hierzu Informationen aus verschiedenen strukturierten und unstrukturierten Quellen, z. B. Twitter, Facebook, YouTube oder Bloomberg, neben Daten, die direkt durch das Unternehmen bereitgestellt werden, sammeln, aufbereiten und analysieren. Die Nutzung von internen und externen Daten beugt dabei möglichen Fehlurteilen aufgrund einer falschen Selbstdarstellung durch das Unternehmen vor. Durch die Anwendung von künstlichen neuronalen Netzen ist dabei keine Beschränkung auf Textdokumente notwendig, sondern es kann auch Bild, Ton- und Videomaterial systematisch analysiert werden. Deep Learning ermöglicht dabei nicht nur die Aufbereitung, sondern auch die zeitlich abhängige Verknüpfung der Daten (Richins et al. 2017). Beispielsweise könnten so Unternehmensinformationen über neue Produkte mit Produktrezensionen verknüpft werden, um Indikatoren über die Produktqualität abzuleiten. Hieraus lassen sich erste Indikationen über, z. B. den Bedarf an Garantierückstellungen ableiten. Ebenfalls kann durch eine explorative Recherche nach Themen wie ”Betrug” oder ”Manipulation” das Prüfungsrisiko, ex-ante, besser eingeschätzt werden. Diese Tätigkeit wäre für menschliche Anwender kaum umsetzbar. Die reine Datenmenge könnte leicht zu einem Information-Overload führen und der dafür notwendige Zeitaufwand wäre unrealistisch (Brown-Liburd & Vasarhelyi 2015).

Ausgehend von den gesammelten Informationen und einer ex-ante Einschätzung des Prüfungsrisikos können auch bedarfsgerechtere Vertragsangebote, d.h. eine genauere Planung der Prüfungsdauer und Prüfungshonorare, unterbreitet werden. Vergleichsgrundlage wären dann nicht mehr nur Unternehmen vergleichbarer Größe und Branche, sondern auch Unternehmen mit einer vergleichbaren ex-ante Risikoeinschätzung, die bereits zuvor von der Wirtschaftsprüfungsgesellschaft geprüft wurden (Issa et al. 2016). Dadurch ist es ebenfalls möglich, den Personalbedarf besser zu planen und Personalkapazitäten effizienter auf Projekte zu verteilen. Abschließend sei noch anzumerken, dass diese Systeme so gestaltet werden können, dass sich diese durch Datenströme[13] weiterentwickeln. Das bedeutet, dass sich die Qualität der Auswertung im Zeitablauf, durch den immer weiterwachsenden Bestand an Daten die bereits ausgewertet wurden, verbessert (Gaber et al. 2005). Dadurch lässt sich die Problematik von Wissensverlust reduzieren, wenn erfahrene Wirtschaftsprüfer eine Wirtschaftsprüfungsgesellschaft verlassen.[14] Unter anderem kann die Wirtschaftsprüfungsgesellschaft so auch auf sich ändernde Umwelteinflüsse zeitnah reagieren.

5.2.5 Identifikation von Fehlerrisiken und interne Kontrollen

Die Identifikation von Fehlerrisiken sowie die Analyse der internen Kontrollen können ebenfalls durch die Automatisierung von Prozessen und Big Data Technologien profitieren, stellen Wirtschaftsprüfer aber auch vor neue Herausforderungen. Durch die Digitalisierung rechnungslegungsbezogener Prozesse und der Speicherung von Daten in Clouds sind Fehlerrisiken zunehmend systematischer Natur und betreffen insbesondere das Thema Datenintegrität (Rega & Teipel 2016). Auf Grund von z. B. Hackerangriffen hat das American Institute of Certified Public Accountants 2016 entsprechend einen Leitfaden zum Thema Cybersecurity veröffentlicht (AICPA 2016). Dieser

[11] Da im Folgenden die Illustration möglicher Potenziale im Fokus steht, werden mögliche rechtliche Beschränkungen bewusst nicht thematisiert.

[12] Beispielsweise bietet IBM mit dem System Watson eine Plattform für solche Lösungen (Melendez 2016).

[13] Ein Datenstrom liefert kontinuierlich neue Daten, z. B. Überwachungskameras die rund um die Uhr aktiviert sind.

[14] Zur Problematik des brain drain, vgl. Knechel et al. (2019).

Leitfaden soll Unternehmen beim Umgang mit solchen Risiken unterstützen und Hinweise zur Kommunikation möglicher Sicherheitsbrüche geben (No & Vasarhelyi 2017).

Aufgrund der andauernden Gefahr von Sicherheitsbrüchen muss sich der Prozess der Risikoidentifikation notwendigerweise zu einem dauerhaften Überwachungsprozess entwickeln. Ansätze im Kontext der kontinuierlichen Prüfung (continuous auditing) zielen daher darauf ab Mechanismen zur Echtzeitüberwachung von rechnungslegungsrelevanten Prozessen zu etablieren (Kiesow & Thomas 2016, Issa et al. 2016). Dadurch könnten bereits unterjährig etwaige Probleme identifiziert und bis zur eigentlichen Abschlussprüfung gelöst werden. Wirtschaftsprüfer würden dann – wie in einer Kommandozentrale eines Kraftwerks – auf Warnsignale der Systeme achten und erst beim Eintreten entsprechender Auffälligkeiten aktiv werden (Dai & Vasarhelyi 2016).[15] Lediglich auf außergewöhnliche Vorfälle im ansonsten transparenten Unternehmen zu reagieren, sogenanntes Audit-by-Excepetion, verteilt den Prüfungsaufwand über das Jahr und reduziert somit den Personalbedarf an Wirtschaftsprüfern und Prüfungsassistenten für die eigentliche Abschlussprüfung (Vasarhelyi et al. 2010).

Demnach würde in einem optimalen System ein Prüfer überhaupt nur dann aktiv werden, wenn eine Auffälligkeit sichtbar wird. Auf Basis der Anfälligkeit der verschiedenen Systeme lässt sich dann auch eine Priorisierung für spätere Prüfungshandlungen ableiten. Stichprobenprüfungen könnten dann stark reduziert auf maschinell identifizierte Risikogruppen angewandt und eine Abschlussprüfung größtenteils fließend durchgeführt werden. Dies setzt allerdings eine nahtlose Vernetzung zwischen Systemen der Wirtschaftsprüfungsgesellschaft und denen der Mandanten voraus (Kiesow & Thomas 2016).

Die Dokumentation des Verständnisses der rechnungslegungsrelevanten Prozesse und Kontrollen kann ebenfalls ressourcenschonend und somit effizienter gestaltet werden. Während klassischerweise vor Ort Besichtigungen durchgeführt werden, könnten hierfür auch Drohnen, Satellitenbilder oder Videos eingesetzt werden (Appelbaum & Nehmer 2017, Brown-Liburd & Vasarhelyi 2015). Auf Grund der geringen Größe und leichten Bedienbarkeit eignen sich Drohnen nicht nur zur mobilen Videoüberwachung in weitläufigen Außenarealen, sondern auch in engen Lagerhallen. Dadurch können auch schwer zugänglich Bereiche untersucht und mögliche Beschädigungen oder Fehlbestände leichter entdeckt werden. Darüber hinaus können, unter Verwendung von 3D-Technologie (z. B. Laserscanning), Lagerbestände genau dokumentiert werden, ohne diese explizit in Augenschein zu nehmen (PwC 2016). Kombiniert mit Artificial Intelligence ließen sich so vollautomatische Systeme zur Dokumentation und Überwachung von Prozessen etablieren. Beispielsweise können Videos von Lagerbeständen mit online verfügbaren Abbildungen von typischen Produkten eines Unternehmens abgeglichen werden, um mögliche Anomalien zu entdecken. Dies lässt sich in vielfältiger Weise auch auf andere Prozesse übertragen, die körperliche Anwesenheit von Mitarbeitern wäre somit nur noch in Ausnahmefällen nötig.

Prüferische Reaktion

Die vielleicht zentralste Änderung könnte sich im Rahmen der eigentlichen prüferischen Reaktionen ergeben. Wie bereits aufgezeigt, basiert die Abschlussprüfung auf dem risikoorientierten Prüfungsansatz und damit verbunden, oft, stichprobenartigen Prüfungen. Dieser Ansatz könnte einer Vollauswertung aller vorhandenen Sachverhalte und somit einem fundamentalen Paradigmenwechsel weichen (Brown-Liburd & Vasarhelyi 2015).

Die möglichen Vorteile einer Voll- gegenüber einer Stichprobenauswertung liegen schon im Namen des Ansatzes begründet. Durch eine Vollauswertung aller Sachverhalte ließe sich, zumindest in der Theorie, die Qualität der Abschlussprüfung steigern (Kiesow & Thomas 2016).[16] Um alle Transaktionsdaten eines Unternehmens effizient prüfen zu können sind allerdings alternative Formen datenanalytischer Verfahren (Audit Data Analytics) notwendig. Beispielsweise zielen Data Mining Verfahren darauf ab einen Datensatz, z. B. alle Buchungen auf ein bestimmtes Sachkonto, auf Muster und mögliche Auffälligkeiten zu untersuchen. Diese Verfahren sind keine neuen Errungenschaften, die nachfolgend dargestellten Potenziale im Kontext von Big Data allerdings schon.

[15] An dieser Stelle sei angemerkt, dass diese Tätigkeit auch in das Aufgabengebiet der internen Revision fallen könnte. Der Bezug zur Wirtschaftsprüfung ergibt primär durch deren Fokus auf die Sicherstellung der Ordnungsmäßigkeit der Rechnungslegung.

[16] Ob dies in Bezug auf mögliche Haftungsfragen von Seiten der Wirtschaftsprüfer überhaupt gewünscht ist, sei hier dahingestellt. Wir verweisen hierzu auf die Ausführungen am Ende des Abschnittes.

Insbesondere die Prüfung von strukturierten und unstrukturierten Daten schafft verschiedenste neue Anwendungsmöglichkeiten (Hayashi 2014, Richins et al. 2017, IAASB/IFAC 2016).

Im Bereich des Journal Entry Testing[17] könnten die Buchungen und deren dazugehöriger Beleg automatisch abgeglichen werden. Um dies zu ermöglichen, existiert in Deutschland bereits eine Initiative zur Erweiterung des PDF-Formats, um strukturierte XML-Daten anzuhängen, die eine maschinelle Weiterverarbeitung wesentlich erleichtern.[18] Auch hier wäre zu prüfen, ob die auf dem Dokument dargestellten Daten mit den versteckten Metadaten übereinstimmen. Trotzdem sind maschinelle Prüfsysteme nicht unfehlbar, sondern lassen sich durchaus beeinflussen (Nguyen et al. 2015).[19] Somit wird es unerlässlich bleiben, künstliche Intelligenzen und deren Entscheidungen regelmäßig zu prüfen und vor allem zu hinterfragen.

Ein weiteres Beispiel sind explorative Datenanalysen. Dabei wird, mit statistischen Methoden, einer oder mehrere Datensätze analysiert um mögliche Zusammenhänge in der Datenstruktur aufzudecken. Die Besonderheit an diesen Verfahren ist, dass ex-ante keine expliziten Hypothesen oder Muster definiert werden.[20] Die Ergebnisse werden folglich auch nicht durch vorher gesetzte Annahmen beeinflusst und werden durch die Nutzung aller verfügbaren Daten auch nicht eingeschränkt (Hunton & Rose 2010, Ruhnke 2017). Primär limitierender Faktor ist hier die Rechenleistung der verwendeten Hardwarekomponenten.

Allerdings bringen solche Vollauswertungen auch Risiken mit sich, welche in der Natur der Daten und der Verfahren begründet sind. Im Gegensatz zu klassischen Prüfungsnachweisen liegen Risiken bei Big Data in der Verlässlichkeit, Herkunft, Vollständigkeit und Interpretationsfähigkeit der Daten. Zwar lässt sich durch die Auswertung digitaler Daten die Quantität verfügbarer Prüfungsnachweise erhöhen, ob diese Informationen in qualitativer Hinsicht einen positiven Mehrwert schaffen, muss zumindest kritisch betrachtet werden (Odenthal 2017, Zhang et al. 2015). Aufgrund der Masse an öffentlich verfügbaren Informationen, deren Herkunft und Qualität in Frage zu stellen ist, stehen Wirtschafsprüfer letztendlich wieder vor einem Selektionsproblem, da bei der Erhebung digitaler Daten bereits eine Abstraktion unternommen wird. Es kann nicht garantiert werden, dass ein digitaler Datensatz alle realen Prozesse im Unternehmen wahrheitsgemäß widerspiegeln kann. Weiterhin lässt sich die Verlässlichkeit der Informationen oft nicht nachprüfen, wenn Informationen anonym verbreitet werden oder die Informationen nicht repräsentativ sind. Entsprechend bleibt der Wahrheitsgehalt bestimmter Informationen zwangsläufig unklar. Weiterhin besteht das Risiko von widersprüchlichen oder nicht eindeutigen Aussagen. Eine binäre (richtig oder falsch) Einteilung würde somit erschwert werden. Für den Einsatz im Rahmen der Abschlussprüfung muss also zunächst geprüft werden, inwieweit solche Informationen überhaupt als Prüfungsnachweise geeignet sind.

In Bezug auf die Anwendbarkeit von datenanalytischen Verfahren, wie zum Beispiel der Regressionsanalyse oder der automatischen Mustererkennung, besteht zudem ein Risiko von fehlerhaften Beurteilungen. Beispielsweise wird bei der Regressionsanalyse eine kausale Beziehung zwischen einer oder mehreren erklärenden Variable(n) und einer abhängigen Variablen unterstellt. In wie weit sich komplexe Unternehmensstrukturen aber adäquat durch mathematische Gleichungen beschreiben lassen mag durchaus kritisch betrachtet werden (Odenthal 2017). Bestenfalls erscheint eine grobe Approximation realistisch. Als Folge dessen kann es dazu kommen, dass Sachverhalte als auffällig oder fehlerhaft eingestuft werden obwohl sie korrekt sind (hoher Alpha-Fehler), beziehungsweise, dass Sachverhalte als unauffällig oder korrekt eingestuft werden obwohl sie fehlerhaft sind (hoher Beta-Fehler). Je nach Modellgüte ergibt sich also ein erheblicher Aufwand zur Validierung der Befunde und möglicherweise sogar ein höherer Aufwand als bei klassischen Stichprobenprüfungen (Ruhnke 2017, Yoon et al. 2015). Eine vergleichbare Problematik ergibt sich beim Thema Mustererkennung, unabhängig ob explorativ oder zum Testen vorformulierter Hypothesen. Zwar wurden in der Literatur gewisse Zahlenmuster als Indikatoren für Auffälligkeiten etabliert (z. B. Benford's Law), aber selbst diese Muster bieten bestenfalls Indikationen für mögliche Fehler. Durch die Einzig-

[17] Die Prüfung des Hauptbuchs eines Unternehmens inklusive Prüfung ausgewählter Belege.

[18] Das ZUGFeRD (Zentraler User Guide des Forums elektronische Rechnung Deutschland) ist eine seit 2014 veröffentlichte Spezifikation zum Austausch elektronischer Rechnungen.

[19] Für ein Anwendungsbeispiel zur Anfälligkeit von Prüfsystemen am Beispiel eines künstlichen neuronalen Netzes zur Bilderkennung siehe Nguyen et al. (2015).

[20] Im Gegensatz dazu werden konventionelle Prüfungshandlungen zur Bestätigung vorformulierter Erwartungen/Hypothesen eingesetzt.

artigkeit unternehmerischer Transaktionen muss im Einzelfall quasi jede auffällige Transaktion validiert werden (Odenthal 2017).[21]

In Bezug auf Einzelfallprüfungen ergeben sich Vorteile durch die Digitalisierung und Vernetzung von Prozessen. Gerade in produzierenden Betrieben erlaubt dies eine Produkt- und Komponentennachverfolgung auf Basis von Technologien wie Barcodes und Radio-Frequency Identification (RFID)-Chips. Möchte ein Wirtschaftsprüfer beispielsweise einzelne Transaktionen nachprüfen, müsste er nicht mehr verschiedene Abteilungen konsultieren, sondern kann sich die Digitalisierung der unternehmensinternen Prozesse zu Nutze machen. Durch eine direkte Verknüpfung von Online-Shop und Lager kann der Eingang einer Bestellung und die Zuordnung zu einzelnen Produkten überprüft werden. Ausgehend von diesem Punkt lassen sich dann alle Bewegungen des Produktes innerhalb eines Unternehmens dokumentieren und der Warenausgang könnte sogar zusätzlich über Videokameras kontrolliert werden. Ebenfalls können RFID-Chips zur Kontrolle des Lagerbestandes und zur Analyse von Warenbewegungen innerhalb des Unternehmens genutzt werden. Sollten diese Daten zusätzlich noch zentral abgelegt werden, können vorgefertigte Data Mining Modelle die Einzelfallprüfung unterstützen oder sogar ganz übernehmen (Vasarhelyi et al. 2015).

Eine weitere Möglichkeit für innovative Prüfungsansätze ergibt sich für die Kontrolle von Risiken auf Personenebene. Um risikobehaftete Mitarbeiter im Unternehmen zu identifizieren kann eine Überprüfung des E-Mail-Verkehrs und der sozialen Medien durchgeführt werden. Hierzu sind zwei Methoden hervorzuheben: (1) Die Sentimentanalyse, auch Sentiment Detection oder Opinion Mining, ist ein Teilgebiet des NLP (Natural Language Processing), dass nach Stimmungsschwankungen in Textdaten sucht. Ausgehend vom E-Mail-Verkehr im Unternehmen können diese Methoden auf unzufriedene Mitarbeiter hinweisen, welchen ein höheres Fehler- und Betrugsrisiko zuzurechnen ist. Nutzt man zudem Daten aus den sozialen Medien, lässt sich ebenfalls ableiten, welche Mitarbeiter ein positives oder negatives Bild an Empfänger außerhalb des Unternehmens weiterleiten. (2) Die Community Detection versucht anhand von Kommunikationsnetzen (z. B. aus XING oder Facebook) soziale Gruppen zu bilden, welche in sich sehr stark vernetzt sind und gegenüber anderen Gruppen sehr schwach. Resultierende Klassifizierungen lassen sich in Bezug auf die Angestellten eines Unternehmens mit bestehenden Abteilungslisten abgleichen, um beispielsweise Angestellte zu ermitteln, die sehr stark mit Mitarbeitern aus anderen Abteilungen oder Unternehmen vernetzt sind. Solche Knotenpunkte können auf anomale Informationsflüsse hindeuten oder ein Hinweis für versteckte Prozesse sein (Pang & Lee 2008).

Abschließend sei noch angemerkt, dass durch digitale Informationen und innovative Prüfungsmethoden zwar zahlreiche Potenziale bestehen, aber eine absolute Prüfungssicherheit wohl utopisch ist und bleibt. Mit genügend krimineller Energie werden sich neue Möglichkeiten finden lassen, um state-of-the-art Systeme zu umgehen. Beispielweise können Systeme zur Analyse des Emailverkehrs dadurch umgangen werden, dass Informationen nur verbal weitergegeben werden. Ebenso geben RFID-Codes auf Lagerkisten noch keine Sicherheit über den tatsächlichen Inhalt der Kisten. Die Expertise und der detektivische Spürsinn von Wirtschaftsprüfern werden folglich nicht zu ersetzen sein. Zudem ist es auch vorstellbar, dass künstliche Intelligenz missbraucht wird um Unternehmensprozesse zu manipulieren. Bei einem möglichen Angriff könnte künstliche Intelligenz nach Schwachstellen in der IT-Infrastruktur eines Unternehmens suchen und hierbei Strategien[22] entwickeln, an die bisher noch kein menschlicher Angreifer gedacht hat.

Auftragsbeendigung

Auch die Berichterstattung über Prüfungsergebnisse könnte durch Digitalisierung weiterentwickelt werden. Dies betrifft sowohl die externe Berichterstattung in Form des Bestätigungsvermerks als auch die interne Berichterstattung in Form des Prüfungsberichts.[23] In Bezug auf die externe Berichterstattung ist eine Abkehr von der Einteilung in uneingeschränkte und eingeschränkte bzw. versagte Bestätigungsvermerke, hin zu einer kontinuierlichen Bewertungsskala denkbar.

[21] Diese Problematik besteht ebenfalls bei modernen Methoden wie der Cluster-Analyse oder der Social-Network-Analyse.
[22] Dass KI dazu im Stande ist neue Strategien zu entwickeln, zeigten Silver et al. (2016) mittels einer KI zum Spiel Go.
[23] Für Erläuterungen zu den beiden Formen der Berichterstattung vgl. Ewert & Wagenhofer (2015).

Die bisherige Bewertung führt bereits heute regelmäßig zu Fehlinterpretationen durch Abschlussadressaten (sog. Erwartungslücke) (Velte 2017, Liggio 1974). Es wäre denkbar, dass der Bestätigungsvermerk – ähnlich einem Sicherheitszertifikat für Webseiten – bald einen tagesgenauen Sicherheits- und Prüfungsstandard angibt. Dies könnte auch die Aussagekraft der neu einzuführenden Berichterstattung zu wichtigen Prüfungssachverhalten (sog. Key Audit Matters) verbessern. Beispielsweise könnte für einzelne Prüfungsfelder angegeben werden, mit welcher Genauigkeit die zugrundeliegenden Sachverhalte geprüft wurden. Auf diese Weise könnte mehr Transparenz für Aktionäre und andere Stakeholder geschaffen werden. Zudem wäre so eine Abkehr von einer rein retrospektiven Abschlussprüfung möglich.

In Bezug auf den Bestätigungsvermerk könnte dem Wirtschaftsprüfer, durch unterstützende Expertensysteme auf Basis von Cognitive Computing[24], zusätzliche Sicherheit gegeben werden. Diese Systeme können die gewonnenen Ergebnisse, mit dem Pool historischer Prüfungsergebnisse und auch Informationen zu nachträglich aufgedeckten Fehlern, z. B. im Rahmen von DPR/BaFin Enforcement-Verfahren[25], abgleichen, um so das Risiko möglicher Fehlurteile zu reduzieren.

Hinsichtlich der internen Berichterstattung kann durch die Abschlussprüfung auch ein zusätzlicher Mehrwert für Unternehmen geschaffen werden. Bereits heute können über die IT-Systemprüfung nach IDW PS 330 Hinweise zum Ausbau der digitalen Infrastruktur gegeben werden. Durch Analysen der Prozessabläufe können zudem Potenziale zur Effizienzsteigerung aufgezeigt werden.[26]

Abschließend bleibt noch das Thema Dokumentation und Archivierung der Prüfungsdokumente zu klären. Eine zunehmend digitale Archivierung von Prüfungsunterlagen ist eine sicherlich naheliegende Erwartung, insbesondere da für Prüfungshandlungen regelmäßig auf Vorjahresunterlagen zurückgegriffen werden muss. Bei einer ausschließlich digitalen Aufbewahrung der Dokumente ergeben sich aber zusätzliche Sicherheitsanforderungen an die IT-Infrastruktur der Prüfungsgesellschaften sowie die Kompatibilität der IT-Systeme von Wirtschaftsprüfer und Mandant (Rega & Teipel 2016, Kiesow & Thomas 2016). In Bezug auf einen externen Datenzugriff muss die Sicherheit der Daten, zu jeder Zeit, gewährleistet sein. Entsprechend muss nicht nur die Sicherheit der verwendeten (Cloud-)Speicher kritischer hinterfragt werden, sondern auch der physikalische Standort dieser Speicher. Etwaige Sicherheitsbrüche oder der Zugriff durch Serverbetreiber oder externe Institutionen[27] führen nicht nur zu einem Verlust vertraulicher Daten, insbesondere aber wird das Vertrauensverhältnis zwischen Prüfer und geprüftem Unternehmen nachhaltig geschädigt (Zhang et al. 2015, Dai & Vasarhelyi 2016). Gerade bei der Abschlussprüfung spielt die Reputation des Prüfers und das Vertrauen zwischen beiden Seiten eine zentrale Rolle. Wird dieses Vertrauen durch Sicherheitsbedenken beeinträchtigt, kann dies zu einem Verlust von Mandaten und negativen wirtschaftlichen Folgen für Prüfer und Mandant führen.[28]

5.3 Mehrwert der Abschlussprüfung versus Unabhängigkeit des Prüfers

Neben der Abschlussprüfung erbringen Wirtschaftsprüfungsgesellschaften für Mandanten oft noch weitere Dienstleistungen wie Steuer-, Rechts- und andere Beratungsleistungen sowie sonstige Bestätigungsleistungen. Bei den vier größten Gesellschaften machen diese Leistungen, in 2015, 43,4% und bei kleineren Gesellschaften 23,3% der jährlichen Honorare aus (WPK, 2015). Grundsätzlich können bestimmte Beratungsleistungen gut von Wirtschaftsprüfungsgesellschaften erbracht werden, da diese durch die Abschlussprüfung bereits mit dem Unternehmen vertraut sind und angenommen werden kann, dass ihnen, durch ihre Berufspflichten (vgl. § 43 WPO), im Umgang mit sensiblen Daten mehr Vertrauen geschenkt wird als anderen externen Dienstleistern. Problematisch sind hier allerdings Gefährdungen der Unabhängigkeit, wenn beispielsweise selbst implementierte Kontrollsysteme geprüft

[24] Expertensysteme auf Basis von Cognitive Computing (vgl. IBM Watson in High, 2012) verbessern die Recherche von domänenspezifischen Informationen. Im Cognitive Computing werden verschiedene Bereiche der Künstliche Intelligenz (KI) gebündelt, darunter Deep Learning, Natural Language Processing und das dynamische Lernen.

[25] Für weitergehende Erläuterungen zum Enforcement der Rechnungslegung in Deutschland, Hitz et al. (2012).

[26] Wir verweisen auf Kapitel 5.3 für weitergehende Ausführungen zu zusätzlichen Dienstleistungsmöglichkeiten im Zuge der Digitalisierung der Wirtschaftsprüfung.

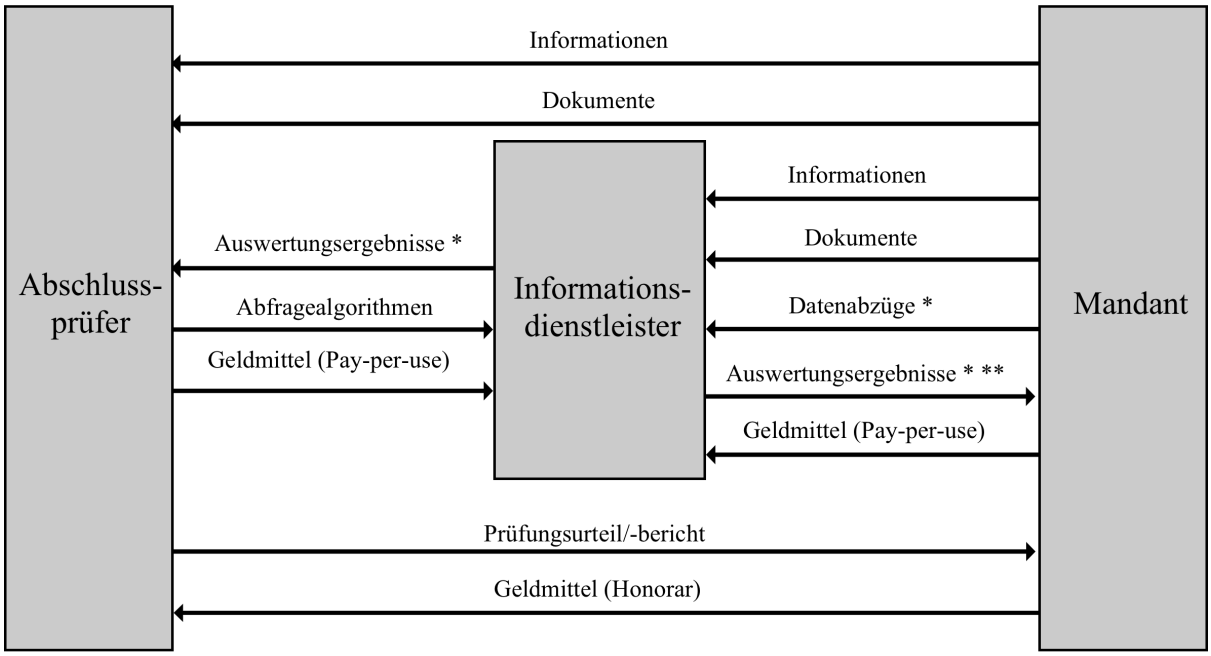
[27] Exemplarisch sei hier der US Patriot Act zu nennen, der amerikanischen Sicherheitsbehörden, in bestimmten Fällen, den Zugriff auf amerikanische Cloud-Speicher ausländischer Unternehmen ermöglicht.

[28] Zur Relevanz von Reputation im Bereich der Wirtschaftsprüfung, vgl. beispielhaft Weber et al. (2008).

werden sollen. Um die Qualität der Abschlussprüfung nicht zu beinträchtigen, ist daher die Art und der Umfang der zulässigen Nichtprüfungsleistungen, nach §§ 319 und 319a HGB, begrenzt beziehungsweise die gleichzeitige Prüfung und Beratung, in bestimmten Fällen, verboten.

Im Zuge der Digitalisierung ergibt sich allerdings zusätzlicher Beratungsbedarf auf Seiten der Unternehmen. Mit einer verstärkt digitalen Abschlussprüfung könnten Wirtschaftsprüfer dabei auch einen Mehrwert für Mandanten schaffen. Neben der IT-Systemprüfung, als Teil der Jahresabschlussprüfung, werden zudem projektbegleitende Prüfungen bei der Einführung von ERP-Systemen (IDW PS 850), IT-Prüfungen außerhalb der Abschlussprüfung (IDW EPS 860), Ausstellung von Softwarebescheinigungen (IDW PS 880), IT Due Dilligence, Prüfung von Dienstleistern bei Outsourcing und Cloud Storage (IDW PS 951 n.F.) oder auch Prüfungen an der Schnittstelle zum Steuerrecht (z. B. GdPdU und E-Bilanz) zunehmend relevanter (Rega & Teipel 2016). Hinzu kommen noch die Potenziale der digitalen Abschlussprüfung. Beispielsweise durch Analysen der IT-Infrastruktur oder die Überwachung von (rechnungslegungsbezogenen) Prozessen können Wirtschaftsprüfer Hinweise für mögliche Effizienzsteigerung aufzeigen (Kiesow & Thomas 2016, Dai & Vasarhelyi 2016). Hieraus ergibt sich dann aber ein möglicher Konflikt im Hinblick auf die Unabhängigkeit des Wirtschaftsprüfers. Da gerade Objekte, die der Wirtschaftsprüfer im Rahmen seiner beratenden Tätigkeit implementiert hat, Gegenstand der Abschlussprüfung wären, kann angenommen werden, dass darin enthaltene Fehler weniger wahrscheinlich aufgedeckt werden (Richter 1977). Aufgrund begrenzter Ressourcen dürfte die Vielzahl der aktuellen IT-Beratungsthemen für kleinere Prüfungsgesellschaften eine Herausforderungen darstellen (Ruhnke 2017). Um trotzdem zusätzliche Beratungsleistungen anbieten zu können, ohne die Unabhängigkeit zu gefährden, ließen sich Intermediäre einsetzen, wie im Folgenden Absatz geschildert wird. Die Wertschöpfungskette im Rahmen der Abschlussprüfung würde dabei um eine dritte Partei erweitert werden. Nach Kiesow & Thomas (2016) können sogenannte Informationsdienstleister eine fachliche Brücke zwischen Wirtschaftsprüfungsgesellschaften und den zu prüfenden Unternehmen schlagen und vermeiden, dass sich jede Prüfungsgesellschaft selbst das teils sehr spezielle IT-Fachwissen aneignen muss.

Das als Audit-as-a-Serivce benannte Modell (vgl. Abbildung 5.2) beschreibt einen kontinuierlichen Prüfprozess in dem Informationsdienstleister als Kompetenzträger für die technische Umsetzung der digitalen Mechanismen dienen. Der Großteil der Daten wird hier zwischen Informationsdienstleister und Unternehmen ausgetauscht und der Abschlussprüfer ermittelt dann durch kundenspezifische Abfragealgorithmen die benötigten Informationen. Hierbei soll auch aktiv die Kommunikation zwischen Unternehmen und Informationsdienstleister angestrebt werden, beispielsweise bei auftretenden Unregelmäßigkeiten als Informationsquelle für die Interne Revision. Somit finanziert sich der Informationsdienstleister nicht nur aus Mitteln des Wirtschaftsprüfers, sondern mitunter auch direkt durch den jeweiligen Mandanten. Dieses Modell könnte mittelständischen Wirtschaftsprüfungsgesellschaften eine Chance geben konkurrenzfähig zu bleiben, da nur für die Nutzung der Data Analytics Werkzeuge und Auswertungen bezahlt werden müsste und keine hohen Investitionen getätigt werden müssten. Gleichzeitig bewahren sich Wirtschaftsprüfer die Unabhängigkeit, da Beratungsleistungen primär durch die Informationsintermediäre erbracht werden. Nichtsdestotrotz erfordert dies die Öffnung des Prüfungsmarktes für neue Akteure, da die Abschlussprüfung nur noch indirekt über den Informationsdienstleister erfolgt und nicht mehr direkt über das zu prüfende Unternehmen (Kiesow & Thomas 2016).



* Kontinuierlich, unterjährig
** Beispielsweise für interne Revision, Berichterstattung an Aufsichtsrat

Abbildung 5.2: *Erweiterte Wertschöpfungskette in der Abschlussprüfung nach (Kiesow & Thomas 2016).*

5.4 Praxisorganisation auf Seiten der Wirtschaftsprüfer

Historisch betrachtet ist die Arbeitsbelastung für Wirtschaftsprüfer zu Beginn des Kalenderjahres am höchsten, da dann die meisten Abschlussprüfungen durchzuführen sind (Rega & Teipel 2016). Hinzu kommen, insbesondere bei kapitalmarktorientierten Unternehmen, noch verkürzte Offenlegungsfristen die die Dauer der Abschlussprüfung begrenzen (§ 325 Abs. 4 HGB). Da im Rest des Kalenderjahres weniger Pflichtprüfungen anfallen, ist die Arbeitsbelastung dort tendenziell niedriger. Somit ist auch der faktische Personalbedarf über das Jahr gesehen unterschiedlich hoch. Darüber hinaus ergibt sich ein erheblicher administrativer Aufwand durch verschiedene Dokumentationspflichten. Beispielhaft sei hier die Dokumentation von geleisteten Arbeitsstunden und Reisekosten zu nennen (Rega & Teipel 2016).

Durch Continuous Auditing und Audit-by-Excepetion wird es höchstwahrscheinlich zu einer Entzerrung der Arbeitsbelastung und somit auch zu Änderungen in der Praxisorganisation kommen. Die bereits dargestellten automatisierten Überwachungsmechanismen werden voraussichtlich zu einer Reduzierung der Belastungsspitzen zum Jahresende/Jahresanfang führen, da Vorprüfungen im bisherigen Umfang nicht mehr nötig sind (Byrnes et al. 2014).[29] Für die Durchführung von Prüfungshandlungen im Rahmen der Abschlussprüfung ist ebenfalls von einem geringeren Arbeitsaufwand auszugehen, da standardisierte Prüfungshandlungen automatisiert erfolgen können (Rega & Teipel 2016, Vasarhelyi et al. 2015). Somit besteht auf Seiten der Wirtschaftsprüfungsgesellschaften insgesamt ein geringer Personalbedarf und damit ein geringer administrativer Aufwand für die Planung und Durchführung von Abschlussprüfungen.

Außerhalb der üblichen Zeitfenster für Abschlussprüfungen ist allerdings eher von einem erhöhten Personalbedarf auszugehen. Durch permanente Überwachungssysteme wird der Personalbedarf stärker durch identifizierte Auffälligkeiten diktiert und liegt somit weniger in der direkten Kontrolle der Wirtschaftsprüfungsgesellschaft. Zudem besteht in solchen Situationen wenig Möglichkeit und Zeit zur Einarbeitung in die Besonderheiten des zu prüfenden Unternehmens. Hierdurch entsteht für die verantwortlichen Prüfer also eine zusätzliche Belastungssituation. Je nachdem bei wie vielen Mandaten zeitgleich Auffälligkeiten festgestellt werden, kann der Personalbedarf daher

[29] Vgl. hierzu die Ausführungen in Kapitel 5.2.4 – Auftragsannahme und Planung.

kurzfristig stark ansteigen. Dieses Phänomen könnte beispielsweise bei externen wirtschaftlichen Shocks auftreten. In solchen Fällen kann es zu Engpässen kommen, wenn die für das Mandat verantwortlichen Mitarbeiter und Teams an mehreren Stellen gleichzeitig benötigt werden. Dann wären eine Priorisierung und eine zeitliche Staffelung der Fälle notwendig. Ob die Digitalisierung also letztendlich zu einem geringeren Personalbedarf und einer besseren Praxisorganisation durch bessere Planbarkeit der Mandate führt, kann deshalb nicht eindeutig beantwortet werden.

5.5 Folgen der Digitalisierung für den Berufsstand

Laut Jahresbericht 2016 der Wirtschaftsprüferkammer gab es zum 1. Januar 2017 14.392 bestellte Wirtschaftsprüfer in Deutschland. Im Verhältnis zu den Vorjahren ist die Anzahl der bestellten Wirtschaftsprüfer damit konstant geblieben (2016: 14.389, 2015: 14.407). Wird die Entwicklung der vergangenen Jahrzehnte betrachtet, ist aber eher eine rückläufige Attraktivität des Berufsstandes der Wirtschaftsprüfer erkennbar (durchschnittliche Wachstumsrate 1995-2005: 4,36 % pro Jahr; durchschnittliche Wachstumsrate 2005-2015: 1,64 % pro Jahr). Weiterhin zeichnet sich eine Verschiebung der demographischen Struktur ab. Während 2005 noch 43,22 % der Wirtschaftsprüfer jünger als 45 Jahre waren, liegt der Anteil dieser Altersgruppe aktuell bei nur noch 26,88 %.[30] Ursächlich für diese Veränderungen sind das öffentliche Image des Berufsstandes und die ungewissen Karrierechancen. Wirtschaftsprüfer werden in der Öffentlichkeit oft als „langweilige Bürokraten“ und „Erbsenzähler“ wahrgenommen. Dies führt zu einem geringeren Interesse am Berufsstand bei Berufseinsteigern (Bravidor & Loy 2017). Die Bestellung als Wirtschaftsprüfer setzt darüber hinaus, neben mehreren Jahren Berufserfahrung, auch den Abschluss eines mehrteiligen Examens voraus. Auf Grund der Komplexität der Themen und der zeitlichen Blockung, bestehen die Examen regelmäßig nur knapp 60% der Teilnehmer (WPK, 2018). Da dieses Examen einen wesentlichen Einfluss auf den langfristigen Werdegang hat, werden Berufseinsteiger oft zusätzlich abgeschreckt oder verlassen die Branche bei Nichtbestehen.

Vor dem Hintergrund der bisherigen Ausführungen mag ein Rückgang der Anzahl der Wirtschaftsprüferinnen und Wirtschaftsprüfer wenig problematisch erscheinen. Allerdings besteht ein Risiko, dass der Rückgang des Personalbestandes verstärkt bei jüngeren, aber möglicherweise technisch versierteren und aufgeschlosseneren, Mitarbeitern erfolgt. Im Zuge der Digitalisierung wäre dies langfristig problematisch.

Im Zuge der Digitalisierung ergeben sich verschiedene Anknüpfungspunkte um einen Imagewandel zu erreichen. Grundsätzlich ermöglicht die Digitalisierung ein flexibleres Arbeitsumfeld. Über entsprechende IT-Schnittstellen und cloudbasierte Speicher kann die Prüfungstätigkeit auch direkt aus dem Büro oder teilweise sogar von zu Hause aus erfolgen. Die regelmäßige Anwesenheit beim zu prüfenden Unternehmen ist somit nicht mehr zwingend erforderlich ((Rega & Teipel 2016, Byrnes et al. 2014). Somit könnten, insbesondere für Berufseinsteiger die eher geografisch gebunden sind oder zumindest eine Tätigkeit mit flexiblen Arbeitsmöglichkeiten anstreben, neue Anreize geschaffen werden. Neben der räumlichen Flexibilität ergeben sich auch Möglichkeiten zur besseren Verteilung der Arbeitslast. Die beständige Kritik der über das Jahr ungleichen Arbeitsbelastung, insbesondere in der Busy Season, und die Kritik der oft repetitiven Tätigkeiten könnten durch Continuous Auditing und Audit-by-Exception adressiert werden (Rega & Teipel 2016). Dadurch könnten Arbeitszeiten besser geplant und die Work-Life Balance verbessert werden. Diese wird auch zunehmend von Absolventen und Berufseinsteigern der *Generation Y* angestrebt (Zemke et al. 2000).

Darüber hinaus werden Wirtschaftsprüfer mittelfristig kaum noch ohne ein gewisses Maß an IT-Kenntnissen, mit wachsendem Fokus auf die Datenanalyse, auskommen (Byrnes et al. 2014).[31] Entsprechend ist es von Bedeutung dieses Thema stärker im Rahmen der Berufsausbildung und der Karriereförderung in den Vordergrund zu rücken (Rega & Teipel 2016, Marten et al. 2017). Klassischerweise ist das Recruiting auf Absolventen aus den Bereichen BWL, Wirtschaftswissenschaften und Jura spezialisiert. Durch die Digitalisierung ergeben sich aber nun auch neue Tätigkeitsfelder, die den Beruf attraktiv für Absolventen aus MINT-Fächern (Mathematik, Informatik,

[30] Alle Berechnungen auf Basis der öffentlich verfügbaren Jahresberichte der WPK. Vgl. hierzu `https://www.wpk.de/oeffentlichkeit/berichte/jahresberichte/`.

[31] Wir weisen darauf hin, dass IT-Kenntnisse eine Ergänzung zu den bisherigen fachlichen Anforderungen darstellen und diese nicht ersetzen (Data Analytics Working Group 2016, Ruhnke 2017).

Naturwissenschaft und Technik) machen. Die IT-Systemprüfung beispielsweise ist zwar heute schon integraler Bestandteil der Abschlussprüfung, in den nächsten Jahren wird dieser aber eine herausgehobene Bedeutung zukommen (Rega & Teipel 2016). Darüber hinaus gilt es datenanalytische Verfahren weiterzuentwickeln und diese auch an gesetzliche Vorgaben anzupassen (Alles 2015). Aktuell sind zahlreiche Wirtschaftsprüfungsgesellschaften bereits in Kooperation mit IT-Unternehmen zur Weiterentwicklung bestehender Prüfungstools. Beispielhaft sei hier die Zusammenarbeit von KPMG und IBM's Watson zur Entwicklung von KI Prüfungstools zu nennen (Melendez 2016).[32] Mittel- und langfristig besteht somit ein erhöhter Bedarf an Berufseinsteigern mit fundierten Kenntnissen in MINT-Fächern. Weiterhin bietet das Institut der Wirtschaftsprüfer e. V. (IDW) seit 2016 die Möglichkeit der Spezialisierung als *IT-Auditor*^{IDW}. Die Zertifizierung ist dabei nicht nur auf Mitarbeiter von Wirtschaftsprüfungsgesellschaften beschränkt, sondern richtet sich explizit an externe Dienstleister die im Rahmen der IT-Systemprüfung tätig sind (IDW 2017). Insgesamt werden als Folge der Digitalisierung für Berufseinsteiger aus MINT-Fächern zusätzliche Anreize zum Berufseinstieg geschaffen.

Abschließend ist noch anzumerken, dass durch die Digitalisierung auch neue Anreize für Berufseinsteiger aus den klassischen Fachdisziplinen geschaffen werden können. Die Zeitersparnis durch den Wegfall standardisierter und tendenziell monotoner Prüfungshandlungen schafft mehr Möglichkeiten, um sich intensiver mit komplexen Bilanzierungssachverhalten sowie der Auslegung von Bilanzierungswahlrechten und Ermessensspielräumen zu befassen. Somit kann die Tätigkeit auch für die Kernzielgruppe wieder attraktiver gestaltet werden.

5.6 Folgen der Digitalisierung für die Marktstruktur

Viele der dargestellten Szenarien sind, aufgrund rechtlicher Einschränkungen, wohl eher mittel- als kurzfristig realisierbar. Was sich aber bereits andeutet, sind mögliche Folgen der Entwicklungen für die Marktstruktur. Gerade kleine und mittelständische Wirtschaftsprüfungsgesellschaften werden, wegen fehlender technischer, personeller und wirtschaftlicher Ressourcen nur wenige Möglichkeiten haben neue Trends voran zu treiben (Ruhnke 2017). Hieraus eine zunehmende Marktdominanz der Big 4 Gesellschaften abzuleiten, erscheint aus verschiedenen Gründen trotzdem voreilig.

Grundlegend muss hinterfragt werden, für welche zu prüfenden Unternehmen eine zunehmende Digitalisierung der Wirtschaftsprüfung überhaupt Relevanz hat. Stark technologisch aufgestellte Unternehmen, welche auch bereits heute schon Technologien wie Blockchain etablieren, sind eine offensichtliche Zielgruppe. Für etwaige Veränderungen der Marktstruktur wird mitentscheidend sein, wie gut neue Prüfungsprozesse in die bestehenden ERP-Systeme der Unternehmen implementierbar sind. Dies ist in der Praxis mit erheblichen Herausforderungen verbunden (Kiesow & Thomas 2016). Eine PwC (2017) Studie mit knapp 100 befragten Unternehmen (39% börsennotiert) zeigt, dass mehrheitlich kein Ausbau oder Wechsel der ERP-Systeme geplant ist. Insoweit stehen Prüfungsgesellschaften also vor der Herausforderung digitale Prüfungsprozesse an verschiedene IT-Infrastrukturen anzupassen. Ähnlich einer Branchenspezialisierung könnten Wirtschaftsprüfungsgesellschaften durch eine Spezialisierung auf bestimmte ERP-Systeme einen Wettbewerbsvorteil realisieren. Die Digitalisierung könnte darüber hinaus auch die Markteintrittsbarrieren für neue Akteure reduzieren und so den Markt kompetitiver gestalten. Theoretisch könnten hoch spezialisierte IT-Unternehmen, wie z. B. Google, bald auch Prüfungsleistungen anbieten (Ruhnke 2017).

5.7 Fazit

Der Berufsstand der Wirtschaftsprüfer blickt, national und international, auf eine reiche und Jahrhunderte überspannende Geschichte zurück. Der Berufsstand hat dabei mehrfach unter Beweis gestellt, dass er sich an strukturelle Veränderungen anpassen kann (Markus 1996). Aufgrund der zunehmenden Bedeutung von Themen wie Artificial Intelligence, Big Data, Cloud Computing und Audit Data Analytics gibt der vorliegende Beitrag einen Überblick über mögliche Chancen und Risiken für die Wirtschaftsprüfung.

[32] Vergleiche M2 Presswire (2016) für einen Anwendungsfall bei PwC. Vergleiche Agnew (2016) für Anwendungsfälle bei EY und Deloitte.

Der Prüfungsprozess könnte prinzipiell auf allen Stufen von einer erhöhten Digitalisierung profitieren. Die möglichen Potenziale sind aber kritisch zu hinterfragen. Zwar ermöglichen automatisierte Vollauswertungen und Methoden aus dem Bereich der Audit Data Analytics die effiziente und umfangreiche Analyse unternehmensspezifischer Daten, allerdings resultiert hieraus nicht zwangsläufig ein höheres Maß an Prüfungssicherheit. Dies liegt vor allem darin begründet, dass automatisierte Auswertungen primär Anomalien aufdecken. Nicht jede Anomalie stellt aber notwendigerweise auch einen expliziten Fehler dar. Eine diskrete Einteilung von Anomalien in richtig oder falsch ist somit schwierig zu realisieren und erfordert zusätzliche Auswertungen. Die Eignung der vielfältigen Datenformate stellt zudem neue Anforderungen an Prüfungsnachweise.

Durch eine starke Digitalisierung der Wirtschaftsprüfung ergeben sich auch Potenziale über den Prüfungsprozess hinaus. Dies liegt darin begründet, dass Fehlerrisiken in digitalen Systemen zunehmend systematischer Natur sind. Das Aufzeigen und Beheben etwaiger Schwachstellen, welche im Rahmen der Abschlussprüfung aufgedeckt werden, wäre daher ein Mehrwert für die Mandanten. Aufgrund der nur begrenzt möglichen simultanen Prüfung und Beratung ist dies allerdings nur eingeschränkt möglich.

Chancen und Risiken der Digitalisierung für Prüfungsgesellschaften und den Berufsstand werden besonders durch geänderte fachliche Anforderungen und den Personalbedarf determiniert. Durch die Nutzung, beziehungsweise Entwicklung, datenanalytischer Verfahren und dem Rückgang tendenziell monotoner Prüfungshandlungen wird der Berufseinstieg nicht nur für Absolventen aus den klassischen Fachbereichen, sondern auch für Absolventen aus den MINT-Fächern, welche verstärkt über IT-Kenntnisse verfügen, attraktiver. Die Folge der Automatisierung von Prüfungshandlungen ist jedoch ein insgesamt geringerer Personalbedarf. Darüber hinaus sind IT-Kenntnisse kein Ersatz für fundierte Kenntnisse im Bereich Rechnungslegung und Wirtschaftsprüfung, sondern eine notwendige Ergänzung dieser.

Eine digitale Revolution in der Wirtschaftsprüfung hätte ebenfalls einen Einfluss auf die Markstruktur. Auf Grund der hohen Investitionskosten für die Entwicklung neuer Verfahren ergeben sich ressourcenbedingte Vorteile für größere Prüfungsgesellschaften. Fraglich ist allerdings, ob auch eine entsprechende Nachfrage auf Seiten der Unternehmen besteht und ob neue Anwendungen auf die verschiedenen IT-Infrastrukturen angewendet werden können. Zudem könnten spezialisierte IT-Dienstleister eine neue Konkurrenzsituation schaffen, sollte die Abschlussprüfung keine Vorbehaltsaufgabe mehr bleiben.

Welche der zahlreichen Entwicklungstrends sich letztendlich durchsetzen werden, bleibt abzuwarten. Menschliche Abschlussprüfer werden auf Grund ihrer Expertise und der Komplexität von Bilanzierungssachverhalten aber auch langfristig kaum zu ersetzen sein.

5.8 Literatur

Agnew, H. (2016), 'Technology transforms big four hiring practices', *Financial Times (online)* .
URL: *https://www.ft.com/content/d5670764-15d2-11e6-b197-a4af20d5575e.*

AICPA (2016), 'Aicpa proposes criteria for cybersecurity risk management'.

Alles, M. G. (2015), 'Drivers of the use and facilitators and obstacles of the evolution of big data by the audit profession', *Accounting Horizons* **29**(2), 439–449.

Appelbaum, D. & Nehmer, R. A. (2017), 'Using drones in internal and external audits: An exploratory framework', *Journal of Emerging Technologies in Accounting* **14**(1), 99–113.

Bartmann, A., Hufgard, S. & Streller, Weltner, V. (2018), Digitaler aufbruch in der wirtschaftsprüfung und beratung, *in* Deggendorfer Forum zur digitalen Datenanalyse e.V., ed., 'Digitalisierung der Prüfung', pp. 25–40.

Bower, J. L. & Christensen, C. M. (1995), 'Disruptive technologies: Catching the wave', *Harvard Business Review* **73**(1), 43–53.

Bravidor, M. & Loy, T. (2017), 'Looking for facts. entwicklung der absolventenzahlen in finanzierung, rechnungswesen und steuerlehre', *Steuer und Wirtschaft* **94**(3), 281–299.

Brösel, G., Freichel, C., Toll, M. & Buchner, R. (2015), *Wirtschaftliches Prüfungswesen*, 3., vollständig überarbeitete auflage edn, München.

Brown-Liburd, H. & Vasarhelyi, M. A. (2015), 'Big data and audit evidence', *Journal of Emerging Technologies in Accounting* **12**(1), 1–16.

Byrnes, P., Criste, T., Stewart, T. & Vasarhelyi, M. (2014), 'Reimagining auditing in a wired world'.

Christensen, C. M. (1997), 'The innovator's dilemma, when new technologies cause great firms to fail', *Harvard Business School Press* .

Curtis, L. H. & Schulman, K. A. (2006), 'Overregulation of health care: Musings on disruptive innovation theory', *Law and Contemporary Problems Distributional Issues in Health Care* **69**(4), 195–206.

Dai, J. & Vasarhelyi, M. A. (2016), 'Imagineering audit 4.0', *Journal of Emerging Technologies in Accounting* **13**(1), 1–15.

Data Analytics Working Group (2016), 'Request for input: Exploring the growing use of technology in the audit, with a focus on data analytics'.
URL: *www.ifac.org*

Ewert, R. & Wagenhofer, A. (2015), *Externe Unternehmensrechnung*, 3., aktualisierte auflage edn, Berlin u.a.

Gaber, M., Zaslavsky, A. & Krishnaswarmy, S. (2005), 'Mining data streams a review', *ACM Sigmod Record* **34**(2), 18–26.

Gerber, T. (2013), 'Xerox' scanner-fehler: Buchstabentausch auch bei höheren qualitätsstufen'.

Goldshteyn, M., Gabriel, A. & Thelen, S. (2013), *Massendatenanalysen in der Jahresabschlussprüfung: Grundlagen und praktische Anwendung mit Hilfe von IDEA*, Düsseldorf.

Grosan, C. & Abraham, A. (2011), 'Intelligent systems', *Intelligent Systems Reference Library* pp. 149–185.

Haas, J. (2017), 'Welchen einfluss wird die künstliche intelligenz in finanz-anwendungen auf die wirtschaftsprüfung nehmen?'.
URL: *http://hwpartners.de/wp-content/uploads/IDW-BDI-digital-summet-10-2017.pdf*

Hayashi, M. (2014), 'Thriving in a big data world', *MIT Sloan Management Review* **55**(2), 35–39.

Hitz, J.-M., Ernstberger, J. & Stich, M. (2012), 'Enforcement of accounting standards in europe: Capital-market-based evidence for the two-tier mechanism in germany', *European Accounting Review* **21**(2), 253–281.

Hunton, J. E. & Rose, J. M. (2010), '21st century auditing: Advancing decision support systems to achieve continuous auditing', *Accounting Horizons* **24**(2), 297–312.

IAASB/IFAC (2016), 'Exploring the growing use of technology in the audit, with a focus on data analytics'.

IDW (2017), 'It-auditor idw'.
URL: *https://www.idw.de/idw/im-fokus/it-auditor-idw*

Issa, H., Sun, T. & Vasarhelyi, M. A. (2016), 'Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation', *Journal of Emerging Technologies in Accounting* **13**(2), 1–20.

Kiesow, A. & Thomas, O. (2016), 'Digitale transformation in der wirtschaftsprüfung, in: Die wirtschaftsprüfung', *Die Wirtschaftsprüfung Supplement* **69**(13), 709–716.

Knechel, R. W., Mao, J., Qi, B. & Zhuang, Z. (2019), 'Is there a brain drain in auditing? the determinants and consequences of auditors' leaving public accounting', *Working Paper* .

Liggio, C. D. (1974), 'The expectation gap: The accountant's legal waterloo?', *Journal of Contemporary Business* **3**(1), 27–44.

Lückmann, R. (2014), 'Wirtschaftsprüfung: Buhlen um mandanten'.

Markus, H. B. (1996), *Der Wirtschaftsprüfer – Entstehung und Entwicklung des Berufes im nationalen und internationalen Bereich*, München.

Marten, K.-U., Czupalla, K. & Harder, R. (2017), 'Digitalisierung in der wirtschaftsprüfung und in der internen revision : Herausforderungen für die aus- und weiterbildung', *Die Wirtschaftsprüfung* **70**(21), 1233–1241.

Melendez, C. (2016), 'Artificial intelligence gets into auditing, what's next?'.
URL: *http://www.infoworld.com/article/3044468/application-development/artificial-intelligence-gets-into-auditing-whats-next.html*

Munir, K. (2012), 'The demise of kodax: Five reasons'.
URL: *https://blogs.wsj.com/source/2012/02/26/the-demise-of-kodak-five-reasons/*

Nguyen, A., Yosinski, Y. & Clune, J. (2015), 'Deep neural networks are easily fooled: High confidence predictions for unrecognizable images', *The IEEE Conference on Computer Vision and Pattern Recognition* pp. 427–436.

No, W. G. & Vasarhelyi, M. A. (2017), 'Cybersecurity and continuous assurance', *Journal of Emerging Technologies in Accounting* **14**(1), 1–12.

Obermaier, R. (2017), Industrie 4.0 als unternehmerische gestaltungsaufgabe: Strategische und operative handlungsfelder für industriebetriebe, *in* R. Obermaier, ed., 'Industrie 4.0 als unternehmerische Gestaltungsaufgabe', Wiesbaden, pp. 3–34.

Odenthal, R. (2017), 'Big data und abschlussprüfung', *Die Wirtschaftsprüfung Supplement* **70**(10), 545–554.

Pang, B. & Lee, L. (2008), 'Opinion mining and sentiment analysis', *Foundations and Trends® in Information Retrieval* **2**(H. 1-2), 1–135.

Pomerol, J.-C. (1997), 'Artificial intelligence and human decision making', *European Journal of Operational Research* **99**(1), 3–25.

PwC (2016), 'Are commercial drones ready for takeoff?'.

PwC (2017), 'Digitale abschlussprüfung, studie zum einsatz von technologie im finanz und rechnungswesen'.

Quick, R. (1996), *Die Risiken der Jahresabschlussprüfung*, Düsseldorf.

Rega, I. & Teipel, G. (2016), 'Digitalisierung in der wirtschaft und im berufsstand', *Die Wirtschaftsprüfung* **68**(1), 39–45.

Richins, G., Stapleton, A., Stratopoulos, T. C. & Wong, C. (2017), 'Big data analytics: Opportunity or threat for the accounting profession', *Journal of Information Systems* **31**(63-79).

Richter, M. (1977), 'Die inkompatibilität von jahresabschlußprüfung und unternehmensberatungen durch wirtschaftsprüfer', *Journal für Betriebswirtschaft* **27**(1), 21–42.

Ruhnke, K. (2017), 'Transformation der abschlussprüfung durch big data analytics', *Die Wirtschaftsprüfung* **70**(8), 422–427.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., an den Driessche, G., Schrittwieser, J., Antonoglou, I. Panneershelvam, V., Lanctot, M., Dielemann, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. & Hassabis, D. (2016), 'Mastering the game of go with deep neural networks and tree search', *Nature* **529**(4), 484–489.

Uygun, Y. & Ilie, M. (2018), Autonomous manufacturing-related procurement in the era of industry 4.0., *in* F. Schupp & H. Wöhner, eds, 'Digitalisierung im Einkauf', pp. 81–97.

Vasarhelyi, M. A., Kogan, A. & Tuttle, B. M. (2015), 'Big data in accounting: An overview', *Accounting Horizons* **29**(2), 381–396.

Vasarhelyi, M., Alles, M. & Williams, K. (2010), *Continuous assurance for the now economy*, Institute of Chartered Accountants in Australia, Sydney.

Velte, P. (2017), 'Ökonomische wirkung der berichterstattung des abschlussprüfers über key audit matters im bestätigungsvermerk', *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung* **17**(10), 434–441.

Weber, J., Willenborg, M. & Zhang, J. (2008), 'Does auditor reputation matter? the case of kpmg germany and comroad ag', *Journal of Accounting Research* **46**(4), 941–972.

WhiteOps & ANA (2017), 'Bot Baseline 2016-2017, Fraud in Digital Advertising', (May), 2004.

Yoon, K., Hoogduin, L. & Zhang, L. (2015), 'Big data as complementary audit evidence', *Accounting Horizons* **29**(2), 431–438.

Zemke, R., Raines, C. & Filipczak, B. (2000), *Generations at Work: Managing the Clash of Veterans, Boomers, Xers and Nexters in Your Workplace*, zweite auflage edn, New York, NY.

Zhang, J., Yang, X. & Appelbaum, D. (2015), 'Towards effective big data analysis in continuous auditing', *Accounting Horizons* **29**(2), 469–476.

Chapter 6

Reflection

At the beginning of my doctoral program, the research in this dissertation was motivated and underpinned by the *TUM School of Management*'s vision:

> *"To become one of Europe's leading management schools at the interface to engineering and science, contributing to solutions for the grand societal challenges."*
> *(TUM School of Management)*

My interpretation of this vision is to try to encourage research that takes insights and knowledge from a diversity of disciplines to work on questions from greater societal interest and importance. In that manner, the vision statement is partly the vision for my dissertation, referring to the interdisciplinary character of the essays.

The overall goal of this dissertation is to advance knowledge about the reliability of organizations under uncertain circumstances. Society has to anticipate that uncertain events might happen more frequently, possibly triggered by growing ecological and economic volatility (Ahir et al. 2018). To still maintain consistent organizational performance, research about reliability in organizations proposes that the organizations – as a product of collective behavior of its members – need to have a specific skillset and that processes show an interconnected structure (Weick et al. 2008). This idea is known as organizational mindfulness and especially appreciated in information systems (Dernbecher et al. 2014, Carlo et al. 2012), studying the relationship between digital technology and their users. The concept of organizational mindfulness proposes processes supporting anticipation and resilience in organizations to stay reliable during uncertain events.

With current methods, it is challenging to validate those processes in organizations, since threatening situations are rare, making it nearly impossible to observe the event in real-time (Sutcliffe et al. 2016). Nevertheless, primarily real-time investigations would be promising since the *intelligent behavior* constructs a reliable organization (Weick & Roberts 1993). However, those processes are poorly recallable by retrospection. I offer an elaboration to find methods to investigate the dynamics between organizational reliability and how it emerges from individual behavior. I do so, by using neuroscientific approaches connecting ideas about an collective mind (Weick & Roberts 1993, Sandelands & Stablein 1987).

First, I address the question about the role of the individual in a reliable organization. Isolating the organization, its members, and the environment those are acting in is challenging for an investigator (Aguinis & Molina-Azorín 2015). I propose simulation experiments, where it is possible to distinguish between characteristics of the organization, the environment, and the individuals (Fischer et al. 2018, 2020). Doing so allows theorizing about unknown connections between the different units of analysis. Further, the effects of concepts like organizational mindfulness are mostly investigable in uncertain events only. Hence, real-time investigations in the field become very challenging. I present two studies to foster this challenge. The simulation experiment studies show how it

is possible to theorize about the reliability of organizations and how the individuals contribute to it depending on their skills and behavior. Little is known about using Evolutionary Algorithms (EAs) to simulate and investigate the reliability of complex systems, such as organizations. That is why the studies primarily identify general effects and influencing variables to build a theoretical foundation for future research.

Second, I develop an approach to answering the question of how to measure the collective mind in organizations. As I identified in Chapter 4, it is impossible to distinguish the body of a collective mind at all. Few is known about the structure of the phenomenon, and a collective mind in organizations would only exist as a dynamic and volatile process of interrelating behavior (Durkheim 1895, Sandelands & Stablein 1987, Wegner 1987). Due to this characteristic supra-individual mental states are not yet observable directly, which produces discussions about its validity: A major criticism against mental states at an organizational level is that the concept does not provide an observable body and is regularly interpreted as a metaphor and anthropomorphism (Jones 1995, Walsh 1995). This discussion impedes current research, which is why I offer a conceptual framework to quantify the supra-individual knowledge structures. Doing so enables us to realize the causal role of individual behavior in the complex interrelating network of the collective mind.

I draw on cognitive science, neuroscientific theory, and computational biology to address this dissertation's main research questions. On the individual and collective levels, I use IIT to measure and analyze the complexity of the individual and collective mind. IIT can measure the quality and quantity of phenomenological experience in any physical system and only requires the fulfillment of its five axioms (Oizumi et al. 2014). In the first two simulation studies, the cognitive abilities of the individual agents have been manipulated as well as the capacity of the organization itself. Further, we examine the organization's reliability under uncertain and novel conditions, after the evolution of collective behavior. In the third essay, I use IIT to develop a measure of the collective mind based on interrelating individual behavior. I define the collective mind so it would be distinct and irreducible to individual knowledge structures. That is how it differs from related concepts like shared ore distributed mind (Hutchins 1989, Hodgkinson & Healey 2008).

Research about high-reliability organizations contain essential ideas for the sustainable strategic development of organizations. Those developments are required to perform a digital transformation in firms (Tabrizi et al. 2019). Auditing is a particularly threatened sector to be affected by this economic and technological change since many of their current work processes of auditors have the potential to be automized. Further, auditors might be forced to change their audit processes and organizational structure to adapt to the digitalized business models of their clients (Byrnes et al. 2014, Rega & Teipel 2016, Vasarhelyi et al. 2015). For instance, the auditor's clients can implement new technology to increase their competitiveness, so audit teams need to improve and expand the integration of members with technological expertise to be able to validate the client's systems. This optimization might break up the hierarchical structure and introduces interdisciplinarity in audit teams. High-Reliability Organizations (HROs) can provide templates for managing such teams and creating the correct communication routines so that auditors can ensure high-quality work for clients whose business models might have become more dynamic and unpredictable (Gebauer 2017).

Another reason for the relevance of HRO practices in auditing is that it is plausible that the auditors' core task, the annual statutory audit, has probably less focus in a digitalized economy since firms might be obliged to implement continuous reporting in future: Hence, the auditor needs to validate the reported figures continuously, too. After implementing continuous auditing, the annual report might be only a summary of the continuous reports and from less relevance (Issa et al. 2016, Kiesow & Thomas 2016). Further, continuous auditing implies that auditors have to monitor the reported values continuously and that they have to react in the case of an exception, such as engineers in the control center of a nuclear power plant. Exceptions that require the intervention of an auditor might be rare in such a scenario. However, if automated systems report an incident, the audit team needs to react immediately to avoid delay or the report of critical, fraudulent, or wrong figures (Dai & Vasarhelyi 2016). Practices of resilience, as suggested by organizational mindfulness, provides processes for situations like that and should inspire auditors in practice while developing a proper digitalization strategy (Carlo et al. 2012, Weick et al. 2008).

In the simulation studies, we realize that the research question unfolded high-complexity. Formulating generalizable findings in simulation experiments is already a challenge, but our simulation uncovered the complexity between the individuals' behavior and its role for organizational goals. Even minor manipulations in the environment, the organization, or the individual can change organizations' reliability. Nevertheless, we can formulate the theoretical suggestions to motivate future research, e.g.: Over-specialization during regular tasks can confuse the individuals during uncertain events, and should therefore be avoided. Further, computer simulations can be used to evolve cooperative behavior naturally, even if not intended. Overall, we suggest that a reliable organization might not be the most efficient or effective but one with the right balance between specialization and generalization.

Building on the sociological literature about the idea of a collective mind, I identify strong relations to *connectionism*[1] in cognition. Both argue that the interrelation of micro-mechanisms (e.g., individual behavior or neurons) can create a macro-phenomenon (e.g., the mind or knowledge). Due to this relation, it is possible to align IIT's axioms (Oizumi et al. 2014) with ideas about a supra-individual mind in organizations (Sandelands & Stablein 1987, Weick & Roberts 1993) without violating against existing conceptualizations in organization theory. That enables us to measure the collective mind in organizations, based on the algorithmic framework IIT provides (Albantakis et al. 2014, Mayner et al. 2018), which confirms that the collective mind would be a highly dynamic and volatile phenomenon (Sandelands & Stablein 1987, Walsh 1995) – the mental state would only exist in the process of interrelating behavior of its members.

In general, this dissertation contributes to the literature investigating multiple levels of analysis. I present solutions to investigate the dynamics between the individual level and the organizational level of analysis. This is possible by applying the methods and theories used in computational biology and neuroscience.

The simulation studies provide experiences and insights about investigating the complexity of individual cognitive limits and the organizational structure concerning reliability in uncertain circumstances. Little is known in that scope, which is why our insights provide suggestions for further research. In particular, we showed that the individual's representation of the environment depends on the size of the organization it belongs to, which has an impact on the overall reliability. This hypothesis can help to evaluate and to optimize reliability in research beyond the laboratory. Our studies identify variables that can have significant effects on reliability in organizations, e.g., the organization's size, perceived information, or the frequency of interactions with the group members.

Measuring the collective mind of organizations adds to the literature in organizational studies. We provide advancement for theoretical discussions on the frontier of organizational cognition in general. It is difficult to provide evidence of a collective mind since it lacks a physical body. My framework offers an approach to argue for a collective mind's existence as a result of a dynamic, interrelated process and methods to quantify the complexity of that mind. Doing so can inspire and improve studies about organizational mindfulness by measuring the potential for mindfulness in an organization. The framework also contributes to studies building on the concept of supra-individual knowledge structures. It isolates the organizational knowledge from the knowledge within the individuals' minds (Walsh 1995, Alavi & Leidner 2016), e.g., transactive memory systems (Wegner 1987, Wegner et al. 1985) or organizational memory (Walsh 1995).

In this dissertation, I have addressed various difficulties and knowledge gaps in the literature. However, it is not that these questions have been overlooked for a long time; it is instead not trivial to explore them. For example, field studies investigating multiple levels of analysis are challenging and costly. On the other side, providing conceptual frameworks and theorize based on computer simulations has high potential and helped management research to validate their ideas and understand phenomena (Davis et al. 2007). Nevertheless, providing theoretical models also has limitations, which is why they need to be evaluated and strengthened by empirical studies (Knudsen et al. 2019).

The experiment design of our computer simulations does have no strong relations to real-world situations but enabled us to make transparent observations. I want to suggest elaborating on the complexity of the experimental designs to strengthen the statements of the presented studies. A popular simulation design is implementing a predator-prey scenario (Olson et al. 2016, Miikkulainen et al. 2012). In the context of management and organization

[1] *Connectionism* describes the computational-representational understanding of the mind based on interrelating mechanisms, such as neurons.

studies, I would suggest implementing already elaborated laboratory experiments. It would be nearby to model the *common-target game* (Joyner 1970, Leavitt 1959).[2]

My approach to measuring the collective mind's structure can serve as a bridge to investigate the organizational and individual level of analysis. Still, it is necessary to gain knowledge about the exact processes underlying the organization's dynamic structure (Morgeson & Hofmann 1999). For instance, the existence of a highly complex organizational structure is no cause for high organizational performance. Following this assumption and adapting philosophical thoughts about the function of consciousness (Rosenthal 2008), the question arises if an organization has a function, respective a performance, or if organizational performance is only an aggregation of individual behavior.

However, a theory on supra-individual knowledge structures helps to advance knowledge about micro-foundations in an organization (Barney & Felin 2013, Kozlowski et al. 2013, Kozlowski & Chao 2012). Studies show that there are relations between supporting individual mindfulness practices and the overall performance in HROs (Hales & Chakravorty 2016, Kudesia 2019, Sutcliffe et al. 2016). This evidence motivates studying the relations between how individual behavior changes after implementing mindful training and how it relates to organizational behavior. The conceptual framework in Chapter 4 can support multi-level research to investigate cognitive processes on the organizational level concerning the individual members (Porac & Tschang 2013, Theiner et al. 2010, Theiner & O'Connor 2010).

What is the impact of a single individual on the behavior of an organized group? This dissertation offers potential answers to this question, based on neuroscientific theories. In particular, I work on the relation between the organization and its members as two distinct subjects, which I did by discussing mental states for both entities. Even though it is hard to imagine supra-individual knowledge structures, I can strengthen arguments in favor of using my theoretical work in organizational cognition. This interdisciplinarity approach can help to understand better how we can improve the impact of an individual's behavior as its role in the organization. This research is profoundly relevant since we are facing high uncertainty in society and the risk of higher unemployment due to the organization's digital transformation.

[2] The common-target game is a laboratory experiment which can be used to theorize about the collective mind(fulness) in organizations (Weick & Roberts 1993).

6.1 References

Aguinis, H. & Molina-Azorín, J. F. (2015), 'Using multilevel modeling and mixed methods to make theoretical progress in microfoundations for strategy research', *Strategic Organization* **13**(4), 353–364.
URL: *http://journals.sagepub.com/doi/10.1177/1476127015594622*

Ahir, H., Bloom, N. & Furceri, D. (2018), 'The world uncertainty index', *Available at SSRN 3275033* .

Alavi, M. & Leidner, D. E. (2016), 'Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues', *MIS* **25**(1), 107–136.

Albantakis, L., Hintze, A., Koch, C., Adami, C. & Tononi, G. (2014), 'Evolution of Integrated Causal Structures in Animats Exposed to Environments of Increasing Complexity', *PLoS Computational Biology* **10**(12), e1003966.
URL: *https://dx.plos.org/10.1371/journal.pcbi.1003966*

Barney, J. & Felin, T. (2013), 'What are Microfoundations?', *Academy of Management Review* **27**(2), 138–155.

Byrnes, P., Criste, T., Stewart, T. & Vasarhelyi, M. (2014), 'Reimagining auditing in a wired world'.

Carlo, J. L., Lyytinen, K. & Boland, R. J. (2012), 'Dialectics of Collective Minding: Contradictory Appropriations of Information Technology in a High-Risk Project', *MIS Quarterly* **36**(4), 1081–1108.

Dai, J. & Vasarhelyi, M. A. (2016), 'Imagineering audit 4.0', *Journal of Emerging Technologies in Accounting* **13**(1), 1–15.

Davis, J. P., Eisenhardt, K. M. & Bingham, C. B. (2007), 'Developing Theory Through Simulation Methods', *Academy of Management Review* **32**(2), 480–499.

Dernbecher, S., Risius, M. & Beck, R. (2014), 'Bridging the Gap – Organizational Mindfulness and Mindful Organizing in Mobile Work Environments', *Ecis* pp. 1–16.

Durkheim, E. (1895), *The Rules of Sociological Method*.

Fischer, D., Mostaghim, S. & Albantakis, L. (2018), 'How swarm size during evolution impacts the behavior, generalizability, and brain complexity of animats performing a spatial navigation task', *GECCO 2018* .
URL: *http://dx.doi.org/10.1145/3205455.3205646*

Fischer, D., Mostaghim, S. & Albantakis, L. (2020), 'How cognitive and environmental constraints influence the reliability of simulated animats in groups', *PLOS ONE* **15**(2), e0228879.

Gebauer, A. (2017), *Kollektive Achtsamkeit organisieren*, Schäffer Poeschl.

Hales, D. N. & Chakravorty, S. S. (2016), 'Creating high reliability organizations using mindfulness', *Journal of Business Research* **69**(8), 2873–2881.

Hodgkinson, G. P. & Healey, M. P. (2008), 'Cognition in Organizations', *Annual Review of Psychology* **59**(1), 387–417.
URL: *http://www.annualreviews.org/doi/10.1146/annurev.psych.59.103006.093612*

Hutchins, E. (1989), 'Distributed Cognition', *IESBS Distributed Cognition* pp. 1–10.

Issa, H., Sun, T. & Vasarhelyi, M. A. (2016), 'Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation', *Journal of Emerging Technologies in Accounting* **13**(2), 1–20.

Jones, M. (1995), 'Organisational learning: Collective mind or cognitivist metaphor?', *Accounting, Management and Information Technologies* **5**(1), 61–77.

Joyner, R. C. (1970), *Computer simulation of individual concept learning in the three-person common target game*, Vol. 7.

Kiesow, A. & Thomas, O. (2016), 'Digitale transformation in der wirtschaftsprüfung, in: Die wirtschaftsprüfung', *Die Wirtschaftsprüfung Supplement* **69**(13), 709–716.

Knudsen, T., A. Levinthal, D. & Puranam, P. (2019), 'Editorial: A Model Is a Model', *Strategy Science* **4**(1), 1–3.

Kozlowski, S. W. J. & Chao, G. T. (2012), 'The Dynamics of Emergence: Cognition and Cohesion in Work Teams', *Managerial and Decision Economics* **33**(5-6), 335–354.

Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T. & Kuljanin, G. (2013), 'Advancing Multilevel Research Design: Capturing the Dynamics of Emergence', *Organizational Research Methods* **16**(4), 581–615.

Kudesia, R. S. (2019), 'Mindfulness as metacognitive practice', *Academy of Management Review* **44**(2), 405–423.
URL: *http://journals.aom.org/doi/10.5465/amr.2015.0333*

Leavitt, H. J. (1959), 'Task ordering and organizational development in the common target game', *Behavioral Science* **5**(3), 233–239.

Mayner, W. G. P., Marshall, W., Albantakis, L., Findlay, G., Marchman, R. & Tononi, G. (2018), 'PyPhi: A toolbox for integrated information theory', *PLOS Computational Biology* **14**(7), e1006343.

Miikkulainen, R., Feasley, E., Johnson, L., Karpov, I., Rajagopalan, P., Rawal, A. & Tansey, W. (2012), Multiagent Learning through Neuroevolution, *in* 'Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)', Vol. 7311 LNCS, pp. 24–46.

Morgeson, F. P. & Hofmann, D. A. (1999), 'The Structure and Function of Collective Constructs: Implications for Multilevel Research and Theory Development', *Academy of Management Review* **24**(2), 249–265.

Oizumi, M., Albantakis, L. & Tononi, G. (2014), 'From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0', *PLoS Computational Biology* **10**(5), 1–25.

Olson, R. S., Knoester, D. B. & Adami, C. (2016), 'Evolution of Swarming Behavior is Shaped By How Predators Attack', *Artificial Life* **22**, 317–330.

Porac, J. & Tschang, F. T. (2013), 'Unbounding the Managerial Mind: It's Time to Abandon the Image of Managers As 'Small Brains'', *Journal of Management Inquiry* **22**(2), 250–254.

Rega, I. & Teipel, G. (2016), 'Digitalisierung in der wirtschaft und im berufsstand', *Die Wirtschaftsprüfung* **68**(1), 39–45.

Rosenthal, D. M. (2008), 'Consciousness and its function', *Neuropsychologia* **46**(3), 829–840.

Sandelands, L. E. & Stablein, R. E. (1987), 'The Concept of Organization Mind', *Research in the Sociology of Organizations* **5**, 135–161.

Sutcliffe, K. M., Vogus, T. J. & Dane, E. (2016), 'Mindfulness in Organizations: A Cross-Level Review', *Annual Review of Organizational Psychology and Organizational Behavior* **3**(1), 55–81.

Tabrizi, B., Lam, E., Girard, K. & Irvin, V. (2019), 'Digital Technology is Not About Technology', *Harvard Business Review* pp. 2–7.
URL: *https://hbr.org/2019/03/digital-transformation-is-not-about-technology*

Theiner, G., Allen, C. & Goldstone, R. L. (2010), 'Recognizing group cognition', *Cognitive Systems Research* **11**(4), 378–395.

Theiner, G. & O'Connor, T. (2010), The Emergence of Group Cognition, *in* 'Emergence in science and philosophy', pp. 92–132.

Vasarhelyi, M. A., Kogan, A. & Tuttle, B. M. (2015), 'Big data in accounting: An overview', *Accounting Horizons* **29**(2), 381–396.

Walsh, J. P. (1995), 'Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane', *Organization Science* **6**(3), 280–321.

Wegner, D. M. (1987), Transactive memory: A contemporary analysis of the group mind, *in* 'Theories of group behavior', Springer, New York, chapter 9, pp. 185–208.

Wegner, D. M., Giuliano, T. & Hertel, P. T. (1985), Cognitive Interdependence in Close Relationships, *in* 'Compatible and Incompatible Relationships', Springer New York, New York, NY, chapter 11, pp. 253–276.

Weick, K. E. & Roberts, K. H. (1993), 'Collective Mind in Organizations: Heedful Interrelating on Flight Decks', *Administrative Science Quarterly* **38**(3), 357.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (2008), 'Organizing for High Reliability: Process of Collective Mindfulness', *Crisis Management* **3**, 31–66.

Appendix A

Contribution to articles and working papers

A.1 Essay 1 (Chapter 2)

- I had the initial paper idea. The article was inspired by my master thesis, submitted in May 2017 to the University of Magdeburg. Still, the present article differs in essential aspects and the findings are based on a different data set.
- I was responsible for the data curation, project administration, software development, visualization, and writing of the original draft.
- Designing the research concept and methodology, validating the experiment results, conducting the formal analysis, investigation, and writing the revisions of the essay was an iterative and cooperative process.
- Larissa Albantakis was supervising the research project.

Dominik Fischer (lead author)

Sanaz Mostaghim (co-author)

Larissa Albantakis (co-author)

A.2 Essay 2 (Chapter 3)

- I had the initial paper idea.
- I was responsible for the data curation, project administration, software development, visualization, and writing of the original draft.
- Designing the research concept and methodology, validating the experiment results, conducting the formal analysis, investigation, and writing the revisions of the essay was an iterative and cooperative process.
- Larissa Albantakis was supervising the research project.

Dominik Fischer (lead author)

Sanaz Mostaghim (co-author)

Larissa Albantakis (co-author)

A.3 Essay 3 (Chapter 4)

I had the paper idea. This essay had no other authors and is only listed in this section for the sake of completeness.

Dominik Fischer (lead author)

A.4 Essay 4 (Chapter 5)

- The topic of this essay was proposed by Robert Obermaier, the book's editor.
- The structure and the development of the idea was elaborated in an iterative and collaborative process.
- The focus of my work was connecting computer scientific topics of digitalization with the challenges in the external audit industry.
- Legal discussions were conducted by my co-author.
- The drafts and the revisions were elaborated in an iterative and collaborative process.

Dominik Fischer (lead author)

Benedikt Downar (co-author)