Talking about the likelihood of risks: an agent-based simulation of discussion processes in risk workshops

Clemens Harten and Matthias Meyer
*Institute of Management Accounting and Simulation,
Hamburg University of Technology, Hamburg, Germany, and*

Lucia Bellora-Bienengräber
Department of Accounting, University of Groningen, Groningen, Netherlands

Abstract

Purpose – This paper aims to explore drivers of the effectiveness of risk assessments in risk workshops.

Design/methodology/approach – This study uses an agent-based model to simulate risk assessments in risk workshops. Combining the notions of transactive memory and the ideal speech situation, this study establishes a risk assessment benchmark and then investigates real-world deviations from this benchmark. Specifically, this study models limits to information transfer, incomplete discussions and potentially detrimental group characteristics, as well as interaction patterns.

Findings – First, limits to information transfer among workshop participants can prevent a correct consensus. Second, increasing the required number of stable discussion rounds before an assessment improves the correct assessment of high but not low likelihood risks. Third, while theoretically advantageous group characteristics are associated with the highest assessment correctness for all risks, theoretically detrimental group characteristics are associated with the highest assessment correctness for high likelihood risks. Fourth, prioritizing participants who are particularly concerned about the risk leads to the highest level of correctness.

Originality/value – This study shows that by increasing the duration of simulated risk workshops, the assessments change – as a rule – from underestimating to overestimating risks, unraveling a trade-off for risk workshop facilitators. Methodologically, this approach overcomes limitations of prior research, specifically the lack of an assessment and process benchmark, the inability to disentangle multiple effects and the difficulty of capturing individual cognitive processes.

Keywords Risk assessment, Risk likelihood, Risk workshops, Simulation, Agent-based modeling, Bayesian networks

Paper type Research paper

1. Introduction

This study investigates conditions affecting the effectiveness of risk assessments in risk workshops [1]. Firms constantly adapt and transform themselves in response to potential risks that may threaten their existence. This entails the need to assess risks correctly – a crucial task in firms' enterprise risk management (COSO, 2017). A failure to distinguish between severe and less severe risks can have serious detrimental consequences, even threatening the continuation of operations. However, this assessment is not a trivial task as decision-makers have to rely on their judgment (Mikes, 2009), which is based on information [2] that is often scattered within and beyond the organization (Neef, 2005).

Risk workshops are a frequently used technique to facilitate the aggregation of this distributed information (COSO, 2017) and allow stakeholders to discuss and assess the impact and likelihood of risks (Boholm and Corvellec, 2016). Risk assessment captures the entire process that determines the severity of a risk after it has been identified (COSO, 2017). The severity of a risk encompasses its potential impact and the likelihood of its occurrence. Risk management literature (van Asselt and Renn, 2011; Quail, 2011) suggests that the risk assessment's effectiveness, in terms of correctly assessing the risks and in terms of the time invested to reach a decision, depends on the design and implementation of this dialogue. We investigate risk workshops' design and implementation from the point where the worst credible impact of a certain risk is evident. Accordingly, the assessment focuses on the likelihood of the risk's worst credible impact. Subsequently, to ensure clarity, "high risks" and "low risks" refer to "high likelihood risks" and "low likelihood risks," respectively, and "risk assessment" refers to the "assessment of the likelihood of a risk."

Because of the difficulty of observing organizational and individual cognitive conditions in discussions (instead of merely noting their outcomes) and the fact that a benchmark (i.e. the correct risk assessment and the time required to achieve it) is ex ante absent in most risk assessments (McNamara and Bromiley, 1997), prior research has been unable to systematically disentangle different sources of (in)effective risk assessments and to describe the unfolding of the discussion over time. We address these challenges by theoretically drawing on the idea of transactive memory and Habermas' (1983) notion of the ideal speech situation. We start by suggesting that risk workshops can be conceptualized as transactive memory systems. Such systems are based on the knowledge stored in each individual's memory, the knowledge about the domain of expertise of the other individuals and the communication of this knowledge. Transactive memory systems represent an attempt to use individuals' information by combining their expertise through a discursive process (Wegner, 1987). Thereafter, we draw on Habermas' (1983) characteristics of an ideal discourse – which include free and full access to the discourse, equal opportunities to express attitudes, desires, and needs and the absence of coercion – to define the most suitable, theoretical-likely conditions to achieve the correct assessment with the least effort. Subsequently, we investigate deviations from this ideal speech situation to determine the risk assessment's unfolding under real-world conditions.

We use agent-based modeling (ABM), namely, simulation experiments that allow agents to follow predefined rules when interacting with other agents and with their environment (Wall and Leitner, 2020). In this study, the agents are workshop participants who communicate to assess a specific risk. ABM allows the development of individual knowledge and its group-level combination, as well as related risk assessment outcomes (Secchi, 2015; Wall and Leitner, 2020). Moreover, our simulation experiments provide a correct assessment – labeled the "benchmark assessment" – against which to evaluate the risk assessment outcome (Labro and Vanhoucke, 2007). To define the "benchmark process," that is, the time required to achieve the benchmark assessment, we start by simulating an

ideal speech situation in which all relevant risk information is shared by the participants. Thereafter, we introduce more realistic scenarios representing deviations from the ideal speech situation. Specifically, we consider the effects of limits to the information transfer among participants (i.e. the receiver only partially accepts the sender's argument because of cognitive load, time pressure, or different backgrounds), incomplete discussions (i.e. the introduction of a decision and termination approach, like voting on the risk assessment after a number of discussion rounds, instead of allowing unlimited information sharing), group characteristics (i.e. unequally distributed information, hierarchical differences and the non-recognition of the possessors of expert knowledge) and specificities of the interaction patterns (e.g. prioritizing higher hierarchical positions in the discussion instead of randomly allowing an introduction of assertions).

We find that, under realistic discussion conditions, it is difficult to attain the benchmark assessment. We therefore generate fine-grained insights on the effects of deviations from the ideal speech situation.

- Even though the risk assessment stabilizes when increasing the number of discussion rounds, limits to information transfer can still prevent a correct consensus [3].
- In incomplete discussions, the discussion conditions suit the correct assessment of either low risks or high risks [4]. An increase in the required number of stable discussion rounds before the leader decides on the risk assessment worsens the correct assessment of low risks.
- Deviations from theoretically detrimental group characteristics lead potentially to higher instead of lower levels of correctness.
- Prioritizing participants who are concerned about a certain risk leads to the highest level of risk assessment correctness.

This paper makes a threefold contribution to research and practice. First, whereas prior risk assessment research focused on overall risks (Aven and Zio, 2014), we raise an awareness thereof that by increasing the average duration of all risk workshops, the assessments change from an underestimation to an overestimation of risks. Thus, the increased correctness of high risks' assessment over the discussion time comes at the cost of a gradually reduced correctness of low risks' assessment. Future researchers are encouraged to refine their research questions by distinguishing between the likelihood of the risks they are targeting, while firms are urged to make allowance for longer discussions if they want to avoid misidentifying high risks.

Second, contrary to the intuitive understanding advocated by previous risk management and group discussion literature, we show that – in the context of risk workshops – the individual characteristics of the theoretically ideal speech situation are not as ideal as presumed (Johnson and Pajares, 1996; Sheffield, 2004). For example, in terms of correctness, a decision made by the leader following an own or the majority assessment outperforms the choice made after waiting for the emergence of a consensus. Firms can learn that the workshop's effectiveness is unlikely to increase after simply improving a single design component. Future research should be cautious when using the ideality notion in discursive settings.

Finally, to the best of our knowledge, this study is the first to systematically introduce a benchmark assessment and process in risk assessment investigations. Generally, an objectively correct assessment is seldom available as a benchmark (Bromiley et al., 2014; McNamara and Bromiley, 1997). Instead, we overcome this limitation and also avoid the

commonly used singular focus on the effort required to achieve a risk assessment, by focusing on the decisions' correctness (Chapman, 1998; Heemstra *et al.*, 2003). Moreover, we allow disentangling the effects of distinct deviations from the ideal speech situation; effects that are otherwise only collectively evident in the risk assessment decision (He *et al.*, 2012). While prior studies accounted for organizational effects, like the order in which participants speak (Hiltz *et al.*, 1986), they were generally unable to capture individual information processing, like the individually assigned importance of received information. We model both types of effects.

2. Theoretical background

2.1 Risk assessment in risk workshops

Risk workshops are instances of group discussions, usually moderated by a facilitator, which provide the basis for a decision made by a leader. Relying on a group requires more effort than, for example, directly soliciting a leader's decision. Collectively, however, the group is expected to make better use of its individual members' information than the individuals would do, as the group can profit from their members' diversity by aggregating their information on different domains (LiCalzi and Surucu, 2012; Lu *et al.*, 2012; Stasser and Birchmeier, 2003).

However, risk workshops (and, more generally, group discussions) often fail to provide reliable (risk) assessments (Hunziker, 2019; Stasser and Titus, 1985). Although scattered, the literature provides some explanations of these outcomes. Among others, the detrimental effects are caused by limited information transfer, because of information overload (Paul and Nazareth, 2010) or the diversity of the participants' backgrounds (LiCalzi and Surucu, 2012). Other arguments point at incomplete discussions owing to time constraints (van Knippenberg *et al.*, 2004) or group characteristics like the lack of familiarity with each other's expertise (Moreland and Myaskovsky, 2000). Moreover, the intra-group interaction patterns are deemed relevant (Katzenbach and Smith, 2015). For example, homogeneity and concurrence seeking – the "groupthink" concept (Janis, 1972) – are related to suboptimal group assessment (Schulz-Hardt *et al.*, 2006). A similar effect could arise when participants are unengaged or when they dominate the discussion (Hunziker, 2019; Quail, 2011). While prior experiment-based laboratory studies are clear about the individual drivers of the quality of the discussion's outcome, they are generally unable to capture the (change of) perceptions of the individual participants and the group during a discussion that is simultaneously affected by multiple conditions (Schulz-Hardt *et al.*, 2006) [5]. However, this process perspective explains at which specific stage of the discussion process a particular decision will be made, in turn unraveling the effectiveness achieved under a particular discussion condition (e.g. terminating the discussion after a certain period of time or focusing on specific participants during the discussion). This study contributes to closing the research gap.

2.2 Risk assessment process: ideal conditions and deviations

We merge the cognitive and discursive perspectives. From a cognitive perspective, we frame risk workshops as an example of distributed cognition. Distributed cognition means that groups make use of individuals' knowledge by combining their expertise (Hauke *et al.*, 2018). Specifically, we rely on transactive memory, a mechanism through which risk workshop participants learn about each other's expertise (i.e. participants build transactive memory) and then identify and combine knowledge in a discursive process. In a risk workshop, a partially differentiated transactive memory system progresses toward an integrated system. In a *differentiated* transactive memory participants have fully disjunct areas of expertise (i.e.

expertise is maximally, unevenly distributed), while in an *integrated* transactive memory all participants have the same knowledge (Wegner, 1987). Transactive memory systems have a positive impact on group performance. This impact is more likely to emerge when group members are familiar with each other's expertise and have initially distributed expertise (Lewis, 2004).

While this cognitive perspective of risk assessments focuses on the group's access to individual knowledge through discussion, the discursive perspective complements the cognitive perspective by focusing on the design of this discussion. Habermas (1983), referring to Alexy (1978), describes the conditions of an ideal speech situation that is theoretically suited to reach a true consensus [6]. In an ideal speech situation:

- all participants competent at speaking about the relevant topic are allowed to participate in the discourse; [7]
- all participants have the same chance of participating by speaking, disagreeing and asking and answering questions, and every aspect can be discussed and criticized; and
- all participants engage in the discussion without differences in power or other forms of coercion.

The ideal speech situation is regarded as a normative standard for a discussion of risks (Horlick-Jones *et al.*, 2001) that ensures the proper sharing and use of individual knowledge in the group.

Real-world discussions are limited by constraints that deviate from the aforesaid ideal speech-situation characteristics. Starting with Habermas (1983) and based on Handy (1986) and our summary of the literature on group discussions, we focus on the following four deviations:

- *Limits to information transfer*: To reach true consensus, the speaker and listener need "shared propositional knowledge, and mutual trust in subjective sincerity" (Habermas, 1982, p. 413). A speech act might not fully convince the receiver if these requirements are not met, with the result that the individual's expertise on a certain risk is not fully incorporated into the assessment.
- *Incomplete discussion*: The ideal speech situation is not limited by temporal constraints as "no preliminary opinion [should remain] permanently withdrawn from discussion and criticism" (Habermas, 1989, p. 177). By contrast, the leader must set time limits to each risk in a workshop (Quail, 2011) and must enforce a termination rule, after which the leader decides on the risk assessment.
- *Specific group characteristics*: As there is no limit to a discussion's length in the ideal speech situation, initial differences in the participants' information access can be resolved by successively sharing information. However, if the discussion remains incomplete (i.e. it is ended before arriving at a true consensus), an *unequal distribution of information* among participants may influence the risk assessment proposed by the group. Moreover, while the equal consideration of each participant's arguments forms a core of the ideal speech situation, in real-world situations *hierarchical differences* may influence the acceptance of arguments. Finally, expertise might go unrecognized (i.e. receivers have *no transactive memory*).
- *Specific interaction patterns*: Habermas (1989, p. 177) calls for participants to "have an equal chance to use representative speech acts" and to "have the same opportunity to use regulative speech acts, that is to give orders and to resist, to

allow and prohibit, to make and take promises." However, it is unlikely that participants will be equally prioritized to speak in real-world discussions (Quail, 2011).

In line with the cognitive and discursive components of our theory, we expect that these deviations from the ideal speech conditions will, *ceteris paribus*, reduce the risk assessments' correctness and increase the required number of discussion rounds.
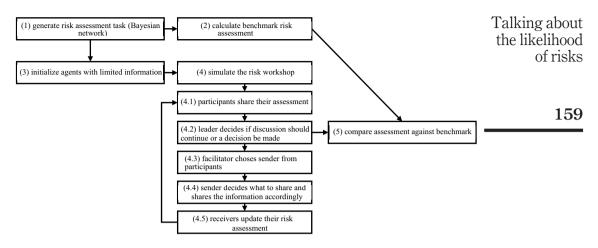
3. Methods

3.1 Overall design

We use a simulation experiment approach, that is, we model the reality of interest with its related processes and outcomes and combine it with an experimental design (Harrison *et al.*, 2007) [8]. First, we align the benchmark process' simulation with Habermas' ideal speech situation. Second, we run four simulation experiments that model the aforesaid deviations from the ideal speech situation to disentangle the extent to which they change the risk assessment's effectiveness. Given the importance of gaining a better understanding of actors' roles in risk management and governance (Hiebl *et al.*, 2018), we model the interaction of risk workshop participants as the exchange of information between agents in an ABM (Lorscheid and Meyer, 2021; Wall and Leitner, 2020).

The risk itself is modeled as a Bayesian network (Fenton and Neil, 2019; Kabir and Papadopoulos, 2019), representing both the discussed risk and the mental model of the participants [9]. Bayesian networks are probabilistic models that describe the conditional probabilities of an event (González-Brenes *et al.*, 2016; Pearl, 2008). Combining ABM and Bayesian networks provides the two components of a transactive memory system, namely, the transactive processes and individual memory systems, reflected, respectively, in the discursive interaction of the ABM's agents and in the likelihood of states represented by Bayesian networks.

3.2 Discussion process and risk assessment model

Each simulation experiment consists of a number of simulation runs. Each simulation run is an entire discussion of a single risk within a risk workshop, and it comprises five stages (Figure 1). The risk structure forms the basis of each discussion (Figure 2). The overall risk assessment (e.g. the likelihood that the introduction of a new product in the market can fail) is derived from the assessment of domain-specific risks (e.g. the likelihood of competitors introducing a similar product or the likelihood that the new product's cost is higher than the customers' willingness to pay). The domain-specific risk assessment is derived from the assessment of issue-specific risks (e.g. the likelihood that productions costs are higher than expected), which, in turn, is rooted in the assessment of specific risk information (e.g. the likelihood that existing machines cannot be adapted to the new product and new machines have to be purchased). The participants' mental model is constructed analogically. The full risk structure contains 40 nodes, comprising 27 information, nine issue and three domain nodes, as well as one node for the overall risk assessment. The individual participants, owing to their diverse backgrounds or priorities, have different risk perceptions (Sjöberg, 2000) and are initially only aware of the existence of the domains and issues related to the information they are provided with during the initialization. Before they can receive information about a certain domain or issue, they have to gain knowledge of this domain or issue's existence by discussing it with other participants [10]. During the discussion, information on the 27 information nodes is exchanged. The other nodes are derived from the state of the information nodes. All nodes are discrete variables in a "low," "medium" or "high" state. Each of these states is assigned a probability that represents the degree of belief that the variable is in a particular state [11]. To reflect

Notes: The figure describes the five stages of each simulation run. (1) A risk assessment task
is randomly generated, according to the Bayesian-network risk structure depicted in Figure 2.
The task is to assess the likelihood of the risk, which is unknown to the workshop participants.
(2) The benchmark assessment is calculated using the complete information from the risk
assessment task. (3) Each participant is provided with some but not all information (i.e. with
limited information), in a way that each information is initially known to at least one
participant. (4) Then, simulation experiments are run, in line with the conditions delineated in
Table 1. The risk is discussed repeatedly and assessed by the group in several discussion
rounds. Specifically, (4.1) all participants share their assessment of the likelihood of the risk;
then (4.2), when certain conditions are met (e.g. the group has reached a consensus), the leader
terminates the discussion and decides on the risk assessment. If the discussion continues, (4.3)
the next participant to share information (i.e. the sender) is chosen and (4.4) shares the
information, (4.5) followed by the other participants (i.e. the receivers) who update their risk
assessment based on this new information. (5) The assessment reached by the workshop is
compared with the benchmark assessment. For example, if a high risk (benchmark assessment)
is assessed to be low, it is a misidentified high risk

Figure 1.
Stages of each
simulation run

a situation where the risk workshop needs to correctly account for a small share of critical
information, we postulate in our Bayesian network that information nodes are individually ten
times more likely to indicate a low than a medium likelihood, and ten times more likely to indicate a
medium than a high likelihood. If a participant believes that a certain information node has a "high"
state (i.e. the state represented by the information node has a high likelihood), the Bayesian network
will reflect this with a higher probability of the corresponding issue-, domain- and overall risk
assessment nodes being in a "high"-risk state. Thus, for the same risk, participants can arrive at
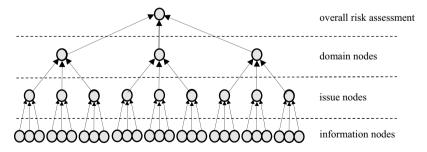different risk assessments, depending on the information available to them.

3.3 Model of the discussion
Nine participants [12] exchange information about the risk at hand. The discussion is
divided into rounds, each comprising a sequence of actions performed by the participants
(see stage 4 in Figure 1). The discussion's outcome is influenced by how it deviates from the
ideal speech situation.

3.3.1 Limits to information transfer. The sender's arguments may not fully convince the receiver. In our model, after receiving information from the sender and when updating their risk assessment, participants will not necessarily fully discard their prior beliefs about the corresponding information node. Instead, a receiver's new assessment of the information node is a weighted average of his/her prior assessment and the sender's assessment [13]. The weight that the receiver attributes to the sender's input differs across receivers and is an aggregate that, in practice, may account for factors like cognitive load, time pressure or the participant's background.

3.3.2 Incomplete discussions. Under real-world conditions, leaders will have to determine the basis on which they will make their assessment decision and when the risk workshop should end. They might rely on their individual risk assessment, on the group consensus or on the majority's assessment. In terms of timing, if the leader adopts a consensual assessment, the discussion could be stopped when the consensus emerges. Otherwise, the leader might stop the discussion if it is not progressing, that is, when the average (numerical) group assessment has stabilized over a certain number of rounds (one, five or ten).

3.3.3 Specific group characteristics. We focus on the impact of three group characteristics.

(1) *Unequal distribution of information*: Participants might not have access to the same amount of information, in which case a larger share of information is provided to some participants.

(2) *Differences in hierarchy*: Information from higher-ranked participants might receive more consideration than information from other participants. Thus, the weight of the information is higher.

(3) *Information about each other's expertise (transactive memory)*: Participants may be unaware of each other's expertise (i.e. receivers lack transactive memory); thus, they cannot differentiate between expert and non-expert senders and will not weigh the information accordingly.

3.3.4 Specific interaction patterns. Risk workshop facilitators decide who is allowed to speak in what order, thereby determining the interaction patterns. Using a random order as a baseline, we investigate the following interaction patterns, giving priority to:

• *Concern*: The probability of being the next sender is higher if the participant's risk assessment is "high."

overall risk assessment

domain nodes

issue nodes

information nodes

Note: The figure represents the risk structure that forms the basis for both the discussion process and each participant's individual mental model

- *Dissent*: The probability of being the next sender is higher if the participant's assessment differs largely from the average (numerical) group risk assessment.
- *Hierarchy*: The probability of being the next sender is higher if the participant is assigned a higher hierarchical position.
- *Homogeneity*: The probability of being the next sender is higher if the participant's risk assessment is close to the average group risk assessment.

4. Results

4.1 Ideal speech-situation conditions

Table 1 provides an overview of our simulation experiments. Figure 3 shows at the top the results of the simulation experiment for the ideal speech-situation conditions. It depicts, per discussion round, the specific proportion of simulated discussions that has reached a particular consensus type or failed to reach a consensus [14]. Before the discussion (i.e. in discussion round zero), no consensus is reached on the risk assessment in 38% of the simulated discussions. The reason is that participants, at the start of the discussion, base their risk assessment only on their limited sets of information. Achieving a (correct) consensus before the discussion is driven by chance.

Moreover, we observe a tendency of initially underestimating risks (i.e. reaching a consensus, but misclassifying high risks). This is because of the lack of knowledge about the existence of certain information nodes. Initially, participants often overlook information about the risk structure and do not account for uncertainty regarding the probabilities of corresponding nodes (i.e. they do not yet know what they do not know). In our model, corresponding to the real-world distribution of risks, most information nodes are in the "low" likelihood state. Consequently, participants underestimate the risk until they, by learning something new about the risk structure, become aware of their – so far unconscious – uncertainty. Therefore, in the early discussion rounds, the low risks are over-proportionally correctly identified, compared to the high risks.

Until discussion round seven, the driven-by-chance consensus drops over all simulated discussions. After this round, an increasing proportion of the discussions results in a consensus – stemming from the increased amount of shared information (thus, from a better knowledge of the risk structure and the corresponding information). After 39 discussion rounds at most, all information is shared and adopted by all participants, resulting in a correct consensus for nearly all discussions [15]. The required maximum of 39 discussion rounds is determined by sum of the 27 information, the nine issue and the three domain nodes that must be shared to attain the overall risk assessment.

Overall, even under ideal speech conditions, it is apparent that a correct group assessment of a risk involves many discussions rounds and is error prone. Moreover, even if the participants reach a consensus, this consensus could be premature and wrong. Hence, the presence of a consensus is only a reliable indicator of a *correct* assessment after a large proportion of information has been shared.

4.2 Simulation experiment 1: limits to information transfer

Figure 3 also shows that, when limiting information transfer, even after 78 discussion rounds – twice as many rounds as in the benchmark process – only 84% of the discussions had reached a correct consensus. As the receivers do not fully integrate new information in their belief updating, senders may have to talk repeatedly about the same information to gradually increase their information's impact on the receivers' risk assessment. At the same
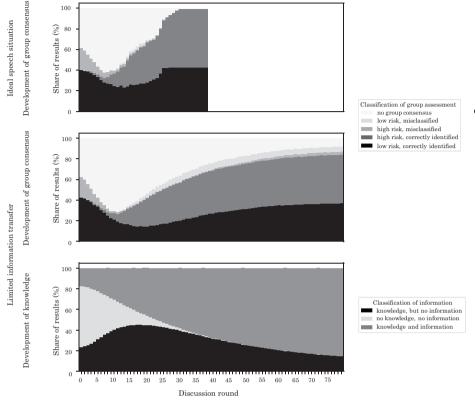
	Ideal speech-situation conditions	Deviations from the ideal speech-situation conditions			
		Simulation experiment 1: Limits to information transfer	Simulation experiment 2: Incomplete discussions	Variations nested under the condition of incomplete discussions	
				Simulation experiment 3: Group characteristics	Simulation experiment 4: Interaction patterns
Experimental conditions					
Receivers retain a part of their prior beliefs	No	**Yes**	Yes	Yes	Yes
Leader's decision approach	Leader follows consensus	Leader follows consensus	**Leader follows: own opinion vs. consensus vs. majority**	Leader follows majority	Leader follows majority
Termination of the discussion	Leader follows consensus	Leader follows consensus	**After one, five or ten stable rounds**[b]	After ten stable rounds	After ten stable rounds
Unequal distribution of information	No	No	No	**Yes/no**	No
Receivers consider hierarchical differences	No	No	No	**Yes/no**	No
Receivers have no transactive memory	No	No	No	**Yes/no**	No
Interaction pattern	Random	Random	Random	Random	**Priority to: concern vs. dissent vs. hierarchy vs. homogeneity**
Outcome variables	% of correct assessments per discussion round	% of correct assessments per discussion round	% of correct assessments, avg. number of discussion rounds	% of correct assessments, avg. number of discussion rounds	% of correct assessments, avg. number of discussion rounds
Number of simulated discussions (n)[a]	1,000	3,768	3,768	7,024	3,200
Number of highflow risks	575/425	2,045/1,723	2,045/1,723	3,858/3,166	1,776/1,424
Results section	4.1	4.2	4.3	4.4	4.5

Table 1.
Overview of the simulation experiments

Notes: The table presents the conducted simulation experiments, along with their respective experimental conditions, outcome variables, number of simulated discussions, number of high and low risks in the benchmark assessment and the section in which the findings are presented. Variables that vary from experiment to experiment are marked in bold. We use a nested design for the simulation experiments targeted at the group characteristics and at the interaction patterns. From simulation experiment 2, we select "leader follows majority" as the decision approach and "after ten stable rounds" as the termination approach, for these experiments. [a]A simulated discussion is the discussion of a single generated risk over several discussion rounds. In each round, a participant shares some information with the group. Each discussion was simulated for 140 discussion rounds, as this was sufficient to reach ten stable rounds for all discussions – which is our strictest stability criterion for the termination of a discussion. Deciding on the number of simulation runs typically involves balancing computational costs and getting representative data generated by the simulation's stochastic process (Lorscheid *et al.*, 2012). [b]A stable round is defined as a discussion round in which the risk assessment does not change from the previous round. A discussion is said to have a number of stable rounds (i.e. the participants' perception is that they do not learn anything more from the discussion) if the average (numerical) group assessment does not differ more than 2% for the same number of consecutive rounds

Note: The graph at the top of this figure depicts the results of the first experiment representing the ideal speech-situation condition (n = 1,000 simulated discussions; high risks: 575 [57.5%]; low risks: 425 [42.5%]), while the middle graph portrays the results of the second simulation experiment (simulation experiment 1) for the limited information transfer condition (n = 3,768 simulated discussions; high risks: 2,045 [54%]; low risks: 1,723 [46%]). Both graphs show, for each discussion round, which proportion of the simulated discussions has reached a particular type of consensus and which proportion has not reached a consensus at this point. The graph at the bottom of this figure depicts the change in the knowledge of the participants during the second simulation experiment. Participants are said to have knowledge about an information node if they include the node in their mental model, and they are said to have information about the node when they receive specific information, either during the initialization or during the discussion

Figure 3.
Development of types
of group consensus
after each discussion
round, under ideal
speech-situation
conditions and with
limited information
transfer

time, as discussion rounds continue, the group assessments' classification becomes stable, sometimes without attaining a correct consensus. Thus, even after many discussion rounds, the unwillingness or inability to fully incorporate the sender's information impedes the achievement of the benchmark assessment.

	Proportion of correct assessments			Avg. number of discussion rounds (%)
	All risks (%)	High risks (%)	Low risks (%)	
Simulation experiment 2: incomplete discussion				
Stop at first group consensus	75.8	57.7	97.2	8.1
One stable round				
Leader follows own opinion	52.5	14.2	98.0	2.1
Leader follows consensus	39.5	0.7	85.5	
Leader follows majority	46.4	1.3	99.9%	
Five stable rounds				
Leader follows own opinion	70.1	65.3	75.9	17.8
Leader follows consensus	39.6	41.0	38.1	
Leader follows majority	70.2	61.2	80.8	
Ten stable rounds				
Leader follows own opinion	77.9	82.6%	72.3	33.5
Leader follows consensus	59.8	68.3	49.7	
Leader follows majority	78.3%	80.9	75.3	
Simulation experiment 3: group characteristics				
Unequal distribution of information with:				
consideration of hierarchical differences, no transactive memory	75.2	91.6%	55.0	34.5
consideration of hierarchical differences, transactive memory	73.4	83.4	61.8	35.5
no consideration of hierarchical differences, no transactive memory	73.1	80.7	63.2	34.6
no consideration of hierarchical differences, transactive memory	75.8	84.4	65.1	36.1
Equal distribution of information with:				
consideration of hierarchical differences, no transactive memory	75.2	82.0	66.3	31.3
consideration of hierarchical differences, transactive memory	77.1	82.8	70.3	31.9
no consideration of hierarchical differences, no transactive memory	73.8	78.2	67.8	31.7

Table 2.
Outcomes of the risk assessment under different deviations

	Proportion of correct assessments			Avg. number of discussion rounds (%)
	All risks (%)	High risks (%)	Low risks (%)	
no consideration of hierarchical differences, transactive memory	78.2%	80.9	75.0%	33.5
Simulation experiment 4: interaction pattern				
Random choice of participants	78.3	80.9	75.3	33.5
Priority given to concerned participants	88.9%	88.9	89.0%	33.8
Priority given to participants with dissenting opinions	79.5	91.2%	64.1	33.3
Priority given to participants with higher hierarchical position	75.2	76.0	74.1	32.2
Priority given to participants close to the group opinion	75.9	70.9	82.7	31.3

Notes: The table depicts the results of the third, fourth and fifth simulation experiment, respectively, and shows the percentage of risks that were correctly assessed, and the average number of discussion rounds before the decision was made. For each experiment, italic values highlight the highest percentage of correct assessments per type of risk and the lowest average number of required discussion rounds. Simulation experiment 2: A stable round is defined as a discussion round in which the risk assessment does not change from the previous round. A discussion is said to have a number of stable rounds (i.e. the participants' perception is that they do not learn anything more from the discussion) if the average (numerical) group assessment does not differ more than 2% for the same number of consecutive rounds. If the leader follows the consensus, but no consensus is reached, the assessment is counted as incorrect. Simulation experiment 3: If the information is unequally distributed, it means that the information is distributed among the participants so that the best-informed participant knows twice as much as the second-best informed participant, who knows twice as much as the least informed participant. If receivers consider hierarchical differences, they weigh the sender's input according to their difference in hierarchy values: $h_{low} = 0.25$, $h_{medium} = 0.5$, $h_{high} = 0.75$. If receivers have no transactive memory, they do not distinguish between the input of an expert sender and a non-expert sender. Simulation experiment 4: When concerned participants are prioritized, the probability of being the next sender is proportional to the probability that they assign the "high risk" state to the overall risk assessment. In a deviation from the standard sequence of actions in the simulation – in this setting – participants select the information to share with a likelihood proportional to the probability they assigned to the "high" state of the respective information node. When dissenting participants are prioritized, the probability of being the next sender is proportional to the difference between their risk assessment and the group's risk assessment. When participants are prioritized based on their hierarchical position, the probability of being the next sender is proportional to a hierarchy factor they are assigned: $h_{low} = 0.25$, $h_{medium} = 0.5$, $h_{high} = 0.75$. When participants close to the group opinion are prioritized, the probability of being the next sender is proportional to the inverse of the difference between their risk assessment and the group risk assessment

Table 2.

4.3 Simulation experiment 2: incomplete discussions
Table 2 aggregates the effects of a leader's three decision approaches, that is, relying on his or her individual risk assessment, accepting the group's consensus or following the majority's opinion. Leaders who follow their own or the majority's opinion outperform the consensus requirement. Regarding all decision approaches, we investigate what happens when the discussion is terminated after one, five or ten stable rounds. We find that this clearly impacts the percentage of correct assessments. Overall risks, a continuation of the discussion generally improves the correctness (e.g. a decision that follows the consensus after ten stable rounds, instead of five stable rounds, improves the overall percentage of correct risk assessments from 39.6% to 59.8%). Intriguingly, correct assessments are different for high and low risks. For example, a comparison of the decision approach with the same number of required stable rounds indicates that the leader will make better decisions by following the majority if the risk is low, but otherwise will improve the decision by relying on his or her individual risk assessment. Terminating the discussion when achieving a first consensus only leads to a correct assessment in 57.7% of the discussions with an actual high risk, while the same termination approach leads to a correct assessment in 97.2% of the discussions with an actual low risk. Moreover, an increase in correctness in high-risk assessments over the discussion time comes at the cost of a slow decrease in correct low-risk assessments. Given that firms want to reduce the severity of the risks that they are facing, and that this severity is the product of the risk's impact and likelihood, *ceteris paribus*, firms will want to identify the high likelihood risks at least correctly and then mitigate these risks. If this holds, based on our findings, firms are encouraged to make allowance for longer discussions to avoid misidentifying high risks.

This trade-off (Figure 3) is partially the result of the previously discussed initial tendency to underestimate risks, as participants – at this point in time – lack knowledge of the complete risk structure, resulting in objectively unjustified certainty ("unknown unknowns"). At this point in time, participants are correct with their "low" assessment, but for the wrong reason. However, as participants subsequently become aware of their lack of knowledge without obtaining information about the likelihood of nodes, they start to overestimate the actual risk as they assign likelihoods to the new nodes. Here, participants also assign small non-zero probabilities to the "medium" and "high" states of the node for the corresponding information node. Consequently, until they learn about the actual state of an increasing number of nodes, many participants assess the overall risk to be high and only switch to a low risk assessment when they learn about the actual state of "low" information nodes.

An increase in the stability requirements is accompanied by an increase in the average number of required discussion rounds. This increase may appear trivial, but it should be noted that it is over-proportional to the number of stable rounds (2.1 for one stable round vs. 17.8 for five stable rounds vs. 33.5 for ten stable rounds). While the overall correct risk assessment only improves in a somewhat linear manner, the time costs of these improvements show a steeper non-linear increase.

4.4 Simulation experiment 3: group characteristics
Table 2 reports the effects of a variation in group characteristics for a condition in which the leader follows the majority after ten stable rounds [16]. As expected, for all risks, we observe the highest correctness (78.2%) when information is equally distributed, receivers do not consider hierarchical differences, and receivers possess transactive memory. Moreover, we find the highest proportion of correctly identified low risks (75.0%) in the same setting. Notably, the highest share of high risks is correctly assessed when there are deviations in all three investigated group characteristics. Under this condition, after the required default of

ten stable rounds required by this simulation experiment, the risk structure has already been learned (i.e. knowledge of the existence of the nodes has been gained); thus, the discussion focuses on the nodes' embedded information. Here, a suboptimal discussion generates noise, as the experts are unable to reduce the other participants' uncertainty. Because not all information is equally discussed, the hierarchically higher participants prevail over the experts, and the expertise of the experts is not recognized. Overall, this does not eliminate the small non-zero probabilities of the "medium" and "high" states of the nodes and leads to an overestimation of all risks. This is the situation in which agents are right with their "high" assessment, but for the wrong reasons [17].

4.5 Simulation experiment 4: interaction patterns
Table 2 indicates that the highest correctness for all risks (88.9%) is observed when prioritizing concerned participants. Prioritizing participants that are close to the group opinion leads to the quickest agreement (31.3 discussion rounds), but at the cost of lower correctness. This aligns with the previous literature's findings that caution against concurrence seeking inherent in the groupthink effect, specifically in risk assessments (Hunziker, 2019; Janis, 1972). Interestingly, we observe improvements when deviating from the equal participation condition suggested by the ideal speech.

5. Contributions and discussion
Our results make three contributions to research and practice. First, we demonstrate that increasing the discussion rounds during a risk workshop may decrease rather than increase the rate of correct assessments for certain risks. Specifically, we identify a potential trade-off between the correct assessment of high and low risks. Along with an increased duration, on average over all risk workshops, the assessments progress from an underestimation to an overestimation of risks. As any improvement of one risk type's correctness reduces the correctness of other types, risk workshop facilitators can choose their discussion termination approaches on this basis. For example, if the correct assessment of high risks is prioritized, attention should be given to the longest possible continuation of the discussion (under the existing resource constraints). We contribute to research by highlighting the peculiarities in the identification of low and high risks over the duration of the group discussions. Future studies should include this distinction in their analyses. For example, would the results of Moreland and Myaskovsky (2000) – who find a positive effect on group performance of a group member's familiarity with the expertise of others – still hold in a risk assessment setting that specifically addresses high or low risks?

Second, we go beyond an ideal speech situation, as we show that this theoretical notion might provide misleading practical guidance. A lengthy discussion that terminates after a large number of stable rounds does not necessarily lead to better outcomes for all risk types. While Stasser and Stewart (1992), following their simulation of political caucuses, concluded that lengthy discussions do not necessarily lead to better decisions, we transfer their finding to a firm-based risk assessment setting, thus indicating that the specific context of discussions is not a boundary condition of this finding. A decision not based on consensual agreement does not prevent good decisions. Thus, we substantiate the conceptual claim that the final risk assessment should be based on the leader's own assessment (Quail, 2011). Rather than allowing everyone to participate in an equal way, we see that facilitators can improve the group's risk assessment by encouraging the participation of those with concerned views. Herewith, we provide evidence supporting the effectiveness of an approach that countervails the concurrence-seeking, groupthink effect in risk workshops. Overall, risk workshop facilitators can learn from our study that an increase in workshop

effectiveness cannot be achieved by simply improving a single design component. Instead, it requires a complete overhaul towards the theoretically ideal conditions, as shown in our benchmark process. Research can profit from our findings by using the identified conditions as a new baseline for further investigations of risk assessments. For example, we complement the work of Katzenbach and Smith (2015) – who argue in favor of determining rules of interactions – by providing evidence of the need to prioritize concerned participants.

Third, we contribute methodologically to the risk assessment literature by introducing a novel approach that uses ABM in combination with simulation experiments. We therefore respond to the call of Bromiley *et al.* (2014), who argue that studies with a known objective risk facilitate an understanding of why and how risk assessments fail to meet expectations. While such a benchmark is usually unavailable in case studies or surveys of the risk assessment practice (McNamara and Bromiley, 1997), it can be generated through a simulation experiment approach. Moreover, in a single study, this approach enables us to disentangle a multitude of effects on the risk assessment. Prior studies mainly focused on either an aggregate effect or on a single effect (Kim and Park, 2010). Finally, it must be noted that our ABM enables us to model individual cognitive processes, including the individual's weighting of the received information and the existence of a transactive memory, and the related group-level outcomes. To the best of our knowledge, this is the first risk assessment study that investigates individual cognitive processes in conjunction with organizational variables. Our modeling might serve as a stepping stone to future risk assessment investigations.

6. Summary and limitations

Risk workshops are a common technique of risk assessment and, if effectively used, constitute a powerful risk management instrument. However, difficulties such as defining benchmarks, disentangling different effects on the risk assessment and capturing individual cognitive processes in discussion processes pose serious challenges to a better understanding of the design and implementation of discussion processes in risk workshops. This study responds to these challenges. It theoretically draws on the notion of transactive memory, links it to the ideal speech conditions, and investigates how deviations from this situation, likely to occur in real-world risk workshops, change the risk assessment outcomes. We ran five simulation experiments rooted in ABM to disentangle the effects of different deviations.

Our results provide fine-grained insights into the processes and outcomes of risk workshops. First, even though the risk assessment *stabilizes* with an increasing number of discussion rounds, limits to information transfer can prevent a *correct* consensus. Second, contrary to our theory and the intuition of group discussion literature, we find that increasing the required number of stable discussion rounds before conducting the risk assessment worsens the correctness for low risks. Third, we show that, for high risks, after ten stable discussion rounds, the co-occurrence of seemingly detrimental group characteristics leads to the highest, instead of the lowest, level of risk assessment correctness. Finally, prioritizing concerned participants, instead of ensuring an equal chance to speak, leads to the highest level of risk assessment correctness.

Admittedly, this paper has limitations that future research should address. First, our analysis simulates a risk workshop that discusses a single risk. While a conscious choice to avoid obfuscating the results with the likely effect of interdependencies across risks, we encourage future studies to use our single risk model as a baseline to investigate these interdependencies' effects. Second, we focus on a classification task that ultimately makes a binary distinction between high and low risks. While we believe that our approach enhances the clarity of the results' communication, future research might be interested in investigating the outcomes of a ternary task. Third, our analysis models nine participants in

the discussion. While nine is within the range of participants common in risk workshops (Ackermann *et al.*, 2014) and the untabulated results of the simulation experiments – ran with three and 18 participants qualitatively support our findings, any related choice is arbitrary; future research should investigate our findings' sensitivity to group size change. Fourth, as we do not address all possible deviations from the ideal speech situation, future research could account for participants' heterogeneous motivation as suggested by Bromiley *et al.* (2014). Likewise, it could clarify factors like hidden agendas or an increase in limits to information transfer over time owing to an increasing instead of a constant cognitive load.

Notes

1. Risk means uncertainty about how potential events may affect the organization. These events may have positive and negative outcomes (COSO, 2017). In this paper, to enhance clarity, we restrict ourselves to the common focus of organizations, that is, those risks that may result in negative outcomes (COSO, 2017). However, our modeling is applicable to both threats and opportunities. For example, when considering interaction patterns in risk workshops, we refer to "concerned" participants; in a threats and opportunities language, a better label would be "concerned or enthusiastic" participants.

2. "Information" refers to the participant's organized data in the context of the risk assessment task, while "knowledge" refers to cognitively processed and aggregated information that enables participants to reach an understanding of the assessed risk.

3. A "correct consensus" refers to a risk assessment that is shared by all the participants of the risk workshop and that corresponds to the benchmark assessment that is ex ante established as correct.

4. We model a risk workshop that deals with a single risk. Investigating potential interdependencies in the risk assessment, when discussing several heterogeneous risks in a single risk workshop, is beyond this study's scope.

5. Usually, laboratory experiment participants are surveyed before and after the discussion. As the capturing (of change) of perceptions during the discussion would disrupt the process, it is generally avoided in laboratory experiments.

6. A consensus is considered true when every competent person agrees with it (Habermas, 1971). Note that a "correct consensus," as previously defined, does not have to be a "true consensus." For example, because the risk workshop conditions allow participants with no knowledge (i.e. not fully competent on the risk considered) to participate, all participants may still reach a correct consensus, albeit not a true consensus.

7. We use the term "discourse," which is common in Habermas' work, as a synonym for "discussion." The latter is used throughout the remainder of this paper.

8. The simulation code and the ODD+D (Overview, Design Concepts and Details + Decision) protocol are available online at www.comses.net. The protocol provides a standard description of ABMs that include human decisions (Grimm *et al.*, 2006; Müller *et al.*, 2013). We use it to detail the information provided in this section.

9. A mental model is an internal representation of a human's understanding of a system (Rouse and Morris, 1986).

10. Gaining knowledge about the existence of a new domain or issue node and obtaining information about the likelihood of this particular node happen in different discussion rounds. When knowledge about the structure of an issue node is acquired, agents simultaneously learn about the existence of the underlying information nodes.

11. For example, a likelihood of 100% for the "low" state signifies that the participant is absolutely certain about the assessment. A likelihood of 80% for the "low" state and, for example, 14% for the "medium" and 6% for the "high" states indicate some uncertainty regarding the actual state of the node.

12. Risk workshops can differ substantially regarding their number of participants. We chose nine participants for our simulation experiment; a group size within the common range for risk workshops (Ackermann *et al.*, 2014).

13. For example, in the ideal speech situation, the non-expert receiver will weigh an expert opinion with 100%. With limited information transfer, a non-expert will weigh an expert opinion with 90% and the prior belief with 10% (e.g. a prior belief of 1% in the high state of an information will turn into a $91\% = 90\% \times 100\% + 10\% \times 1\%$ belief after talking to an expert who assigns 100% to the likelihood of the high state).

14. It is important to note that the Bayesian network is calibrated in a way that always results in a "low risk" or a "high risk" assessment of the overall risk. This simplifies the interpretation of the simulation results. In our Bayesian network, nodes aggregate the input from three other nodes. Because at least some input nodes assign high likelihoods to the "low" or "high" states, as is inevitably the case, the likelihoods assigned to the "medium" state decrease with each level of aggregation. As a result, the participants are presented – *de facto* – with a binary assessment task.

15. Owing to the slight imprecision inherent in the computational framework, 100% correctness is never achieved.

16. For the sake of simplification, a condition for the decision and termination approach must be chosen for both simulation experiments 3 and 4, instead of running simulations experiments for all theoretically possible conditions. We chose the condition that leads to the highest proportion of correct assessments over all risks in the simulation experiments for simulation experiment 2. Untabulated robustness analyses show that changing this condition (e.g. the leader follows consensus after five stable rounds) does not qualitatively change the inferences from our findings.

17. In this simulation experiment, the unequal distribution of information among participants is operationalized so that the best-informed participant, on average, knows twice as much as the second best-informed participant, who in turn knows twice as much as the next best-informed participant, etc. (i.e. we work with a factor of two). In untabulated robustness tests, we run the same simulation experiment with factor 1.5 and factor 2.5. The robustness tests show the same overall direction of effects for all three factors. The only exception is a factor of 2.5, where information is initially highly concentrated within a small group of participants. This increases the participants' initial ignorance of their knowledge ("I don't know what I don't know"), resulting in a slightly lower-level recognition of high risks compared to a setting with a factor of 2.0.

References

Ackermann, F., Howick, S., Quigley, J., Walls, L. and Houghton, T. (2014), "Systemic risk elicitation: using causal maps to engage stakeholders and build a comprehensive view of risks", *European Journal of Operational Research*, Vol. 238 No. 1, pp. 290-299, available at: https://doi.org/10.1016/j.ejor.2014.03.035

Alexy, R. (1978), *Theorie der juristischen Argumentation: Die Theorie des rationalen Diskurses als Theorie der juristischen Begründung*, Suhrkamp, Frankfurt am Main.

Aven, T. and Zio, E. (2014), "Foundational issues in risk assessment and risk management", *Risk Analysis*, Vol. 34 No. 7, pp. 1164-1172, available at: https://doi.org/10.1111/risa.12132

Boholm, Å. and Corvellec, H. (2016), "The role of valuation practices for risk identification", in Power, M. (Ed.), *Riskwork*, Oxford University Press, Oxford, pp. 110-129, available at: https://doi.org/10.1093/acprof:oso/9780198753223.003.0006

Bromiley, P., McShane, M., Nair, A. and Rustambekov, E. (2014), "Enterprise risk management: review, critique, and research directions", *Long Range Planning*, Vol. 48 No. 4, pp. 265-276, available at: https://doi.org/10.1016/j.lrp.2014.07.005

Chapman, R.J. (1998), "The effectiveness of working group risk identification and assessment techniques", *International Journal of Project Management*, Vol. 16 No. 6, pp. 333-343, available at: https://doi.org/10.1016/S0263-7863(98)00015-5

COSO (2017), *Enterprise Risk Management: Aligning Risk with Strategy and Performance*, COSO New York, NY.

Fenton, N.E. and Neil, M. (2019), *Risk Assessment and Decision Analysis with Bayesian Networks*, Second edition., CRC Press, Taylor and Francis Group, Boca Raton.

González-Brenes, J.P., Behrens, J.T., Mislevy, R.J., Levy, R. and DiCerbo, K.E. (2016), "Bayesian networks", in Rupp, A.A. and Leighton, J.P. (Eds), *The Handbook of Cognition and Assessment*, John Wiley and Sons, Hoboken, NJ, pp. 328-353, available at: https://doi.org/10.1002/9781118956588.ch14

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J. and Goss-Custard, J. (2006), "A standard protocol for describing individual-based and agent-based models", *Ecological Modelling*, Vol. 198 No. 1-2, pp. 115-126, available at: https://doi.org/10.1016/j.ecolmodel.2006.04.023

Habermas, J. (1971), "Vorbereitende bemerkungen zu einer theorie der kommunikativen kompetenz", *Theorie Der Gesellschaft Oder Sozialtechnologie – Was Leistet Die Systemforschung?*, Suhrkamp, Frankfurt am Main.

Habermas, J. (1982), *Theorie des kommunikativen Handelns*, 2nd ed., Suhrkamp, Frankfurt am Main.

Habermas, J. (1983), *Moralbewußtsein und kommunikatives Handeln*, Vol. 422 Suhrkamp, Frankfurt am Main.

Habermas, J. (1989), *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*, 3rd ed., Suhrkamp, Frankfurt am Main.

Handy, C.B. (1986), *Understanding Organizations*, Penguin, Harmondsworth.

Harrison, J.R., Lin, Z., Carroll, G.R. and Carley, K.M. (2007), "Simulation modeling in organizational and management research", *Academy of Management Review*, Vol. 32 No. 4, pp. 1229-1245, available at: https://doi.org/10.5465/AMR.2007.26586485

Hauke, J., Lorscheid, I. and Meyer, M. (2018), "Individuals and their interactions in demand planning processes: an agent-based, computational testbed", *International Journal of Production Research*, Vol. 56 No. 13, pp. 4644-4658, available at: https://doi.org/10.1080/00207543.2017.1377356

He, H., Martinsson, P. and Sutter, M. (2012), "Group decision making under risk: an experiment with student couples", *Economics Letters*, Vol. 117 No. 3, pp. 691-693, available at: https://doi.org/10.1016/j.econlet.2011.12.081

Heemstra, F.J., Kusters, R.J. and de Man, H. (2003), "Guidelines for managing bias in project risk management", *International Symposium on Empirical Software Engineering, 2003. ISESE 2003. Proceedings, presented at the 2003 International Symposium on Empirical Software Engineering*, pp. 272-280, available at: https://doi.org/10.1109/ISESE.2003.1237988

Hiebl, M.R.W., Baule, R., Dutzi, A., Stein, V. and Wiedemann, A. (2018), "Guest editorial", *The Journal of Risk Finance*, Vol. 19 No. 4, pp. 318-326, available at: https://doi.org/10.1108/JRF-08-2018-194

Hiltz, S.R., Johnson, K. and Turoff, M. (1986), "Experiments in group decision making communication process and outcome in face-to-face versus computerized conferences", *Human Communication Research*, Vol. 13 No. 2, pp. 225-252, available at: https://doi.org/10.1111/j.1468-2958.1986.tb00104.x

Horlick-Jones, T., Rosenhead, J., Georgiou, I., Ravetzd, J. and Löfstedte, R. (2001), "Decision support for organisational risk management by problem structuring", *Health, Risk and Society*, Vol. 3 No. 2, pp. 141-165. available at: https://doi.org/10.1080/13698570125225.

Hunziker, S. (2019), *Enterprise Risk Management: Modern Approaches to Balancing Risk and Reward*, Gabler Verlag, Wiesbaden, available at: https://doi.org/10.1007/978-3-658-25357-8

Janis, I.L. (1972), *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*, Houghton, Mifflin, Boston.

Johnson, M.J. and Pajares, F. (1996), "When shared decision making works: a 3-year longitudinal study", *American Educational Research Journal*, Vol. 33 No. 3, pp. 599-627, available at: https://doi.org/10.3102/00028312033003599.

Kabir, S. and Papadopoulos, Y. (2019), "Applications of Bayesian networks and petri nets in safety, reliability, and risk assessments: a review", *Safety Science*, Vol. 115, pp. 154-175, available at: https://doi.org/10.1016/j.ssci.2019.02.009.

Katzenbach, J.R. and Smith, D.K. (2015), *Wisdom of Teams: Creating the High-Performance Organization*, Harvard Business Review Press, Boston, MA.

Kim, D.-Y. and Park, J. (2010), "Cultural differences in risk: the group facilitation effect", *Judgment and Decision Making*, Vol. 5 No. 5, pp. 11.

Labro, E. and Vanhoucke, M. (2007), "A simulation analysis of interactions among errors in costing systems", *The Accounting Review*, Vol. 82 No. 4, pp. 939-962.

Lewis, K. (2004), "Knowledge and performance in knowledge-worker teams: a longitudinal study of transactive memory systems", *Management Science*, Vol. 50 No. 11, pp. 1519-1533, available at: https://doi.org/10.1287/mnsc.1040.0257.

LiCalzi, M. and Surucu, O. (2012), "The power of diversity over large solution spaces", *Management Science*, Vol. 58 No. 7, pp. 1408-1421, available at: https://doi.org/10.1287/mnsc.1110.1495.

Lorscheid, I. and Meyer, M. (2021), "Toward a better understanding of team decision processes: combining laboratory experiments with agent-based modeling", *Journal of Business Economics*, available at: https://doi.org/10.1007/s11573-021-01052-x.

Lorscheid, I., Heine, B.-O. and Meyer, M. (2012), "Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments", *Computational and Mathematical Organization Theory*, Vol. 18 No. 1, pp. 22-62, available at: https://doi.org/10.1007/s10588-011-9097-3.

Lu, L., Yuan, Y.C. and McLeod, P.L. (2012), "Twenty-five years of hidden profiles in group decision making: a meta-analysis", *Personality and Social Psychology Review*, Vol. 16 No. 1, pp. 54-75, available at: https://doi.org/10.1177/1088868311417243.

McNamara, G. and Bromiley, P. (1997), "Decision making in an organizational setting: cognitive and organizational influences on risk assessment in commercial lending", *Academy of Management Journal*, Vol. 40 No. 5, pp. 1063-1088, available at: https://doi.org/10.2307/256927.

Mikes, A. (2009), "Risk management and calculative cultures", *Management Accounting Research*, Vol. 20 No. 1, pp. 18-40, available at: https://doi.org/10.1016/j.mar.2008.10.005.

Moreland, R.L. and Myaskovsky, L. (2000), "Exploring the performance benefits of group training: transactive memory or improved communication?", *Organizational Behavior and Human Decision Processes*, Vol. 82 No. 1, pp. 117-133, available at: https://doi.org/10.1006/obhd.2000.2891.

Müller, B., Bohn, F., Dreßler, G., Groeneveld, J., Klassert, C., Martin, R., Schlüter, M., *et al.* (2013), "Describing human decisions in agent-based models - ODD+D, an extension of the ODD protocol", *Environmental Modelling and Software*, available at: https://doi.org/10.1016/j.envsoft.2013.06.003.

Neef, D. (2005), "Managing corporate risk through better knowledge management", *The Learning Organization*, Vol. 12 No. 2, pp. 112-124, available at: https://doi.org/10.1108/09696470510583502.

Paul, S. and Nazareth, D.L. (2010), "Input information complexity, perceived time pressure, and information processing in GSS-based work groups: an experimental investigation using a

decision schema to alleviate information overload conditions", *Decision Support Systems*, Vol. 49 No. 1, pp. 31-40, available at: https://doi.org/10.1016/j.dss.2009.12.007

Pearl, J. (2008), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Rev. 2*, Kaufmann, San Francisco, CA.

Quail, R. (2011), "How to plan and run a risk management workshop", *Enterprise Risk Management*, John Wiley and Sons, Inc., Hoboken, NJ, pp. 155-170, available at: https://doi.org/10.1002/9781118267080.ch10

Rouse, W.B. and Morris, N.M. (1986), "On looking into the black box: prospects and limits in the search for mental models", *Psychological Bulletin*, Vol. 100 No. 3, pp. 349.

Schulz-Hardt, S., Brodbeck, F.C., Mojzisch, A., Kerschreiter, R. and Frey, D. (2006), "Group decision making in hidden profile situations: dissent as a facilitator for decision quality", *Journal of Personality and Social Psychology*, Vol. 91 No. 6, pp. 1080-1093, available at: https://doi.org/10.1037/0022-3514.91.6.1080

Secchi, D. (2015), "A case for agent-based models in organizational behavior and team research", team performance management", *Team Performance Management*, Vol. 21 No. 1/2, pp. 37-50, available at: https://doi.org/10.1108/TPM-12-2014-0063

Sheffield, J. (2004), "The design of GSS-enabled interventions: a Habermasian perspective", *Group Decision and Negotiation*, Vol. 13 No. 5, pp. 415-435, available at: https://doi.org/10.1023/B:GRUP.0000045750.48336.f7.

Sjöberg, L. (2000), "Factors in risk perception", *Risk Analysis*, Vol. 20 No. 1, pp. 1-12, available at: https://doi.org/10.1111/0272-4332.00001

Stasser, G. and Birchmeier, Z. (2003), "Group creativity and collective choice", in Paulus, P.B. and Nijstad, B.A. (Eds), *Group Creativity: Innovation through Collaboration*, Oxford University Press, New York, NY, Oxford, pp. 85-109.

Stasser, G. and Stewart, D. (1992), "Discovery of hidden profiles by decision-making groups: solving a problem versus making a judgment", *Journal of Personality and Social Psychology*, Vol. 63 No. 3, pp. 426-434. available at: https://doi.org/10.1037/0022-3514.63.3.426

Stasser, G. and Titus, W. (1985), "Pooling of unshared information in group decision making: biased information sampling during discussion", *Journal of Personality and Social Psychology*, Vol. 48 No. 6, pp. 1467-1478, available at: https://doi.org/10.1037/0022-3514.48.6.1467

van Asselt, M.B.A. and Renn, O. (2011), "Risk governance", *Journal of Risk Research*, Vol. 14 No. 4, pp. 431-449, available at: https://doi.org/10.1080/13669877.2011.553730

van Knippenberg, D., De Dreu, C.K.W. and Homan, A.C. (2004), "Work group diversity and group performance: an integrative model and research agenda", *Journal of Applied Psychology*, Vol. 89 No. 6, pp. 1008-1022, available at: https://doi.org/10.1037/0021-9010.89.6.1008

Wall, F. and Leitner, S. (2020), "Agent-based computational economics in management accounting research: opportunities and difficulties", *Journal of Management Accounting Research*, available at: https://doi.org/10.2308/JMAR-19-073

Wegner, D.M. (1987), "Transactive memory: a contemporary analysis of the group mind", in Mullen, B. and Goethals, G.R. (Eds), *Theories of Group Behavior*, Springer New York, NY, pp. 185-208, available at: https://doi.org/10.1007/978-1-4612-4634-3_9

Corresponding author
Matthias Meyer can be contacted at: matthias.meyer@tuhh.de