Group Decision and Negotiation (2024) 33:113–145
https://doi.org/10.1007/s10726-023-09856-8

Digital Facilitation of Group Work to Gain Predictable Performance

Henner Gimpel[1,2,3] · Stefanie Lahmer[1,2] · Moritz Wöhl[1,2] · Valerie Graf-Drasch[1,2,3]

Accepted: 19 September 2023 / Published online: 20 October 2023

Abstract

Group work is a commonly used method of working, and the performance of a group can vary depending on the type and structure of the task at hand. Research suggests that groups can exhibit "collective intelligence"—the ability to perform well across tasks—under certain conditions, making group performance somewhat predictable. However, predictability of task performance becomes difficult when a task relies heavily on coordination among group members or is ill-defined. To address this issue, we propose a technical solution in the form of a chatbot providing advice to facilitate group work for more predictable performance. Specifically, we target well-defined, high-coordination tasks. Through experiments with 64 virtual groups performing various tasks and communicating via text-based chat, we found a relationship between the average intelligence of group members and their group performance in such tasks, making performance more predictable. The practical implications of this research are significant, as the assembly of consistently performing groups is an important organizational activity.

Keywords Group work · Consistent group performance · Group support system · Performance prediction · Automated facilitation

1 Introduction

Collaborating in groups to complete work tasks is one of the most established ways of working (Chapman et al. 2006). Hence, consistently (well-) performing groups are of substantial value to companies since group performance becomes somewhat predictable. Predictable performance is desirable in various domains for several reasons. It enables effective planning and efficient working, allowing organizations to plan their operations, projects, and resources more effectively (Albert 2004). When performance is consistent and predictable, estimating timelines, allocating resources, and setting realistic goals becomes easier. This enables better project management,

Extended author information available on the last page of the article reduces the likelihood of delays, and enhances overall productivity. When performance is consistent, it becomes easier to identify and rectify deviations or issues. This reduces the likelihood of errors, defects, or service failures, enhancing overall quality.

The ability of an individual to perform consistently (well) across various types of tasks is determined by one's "individual intelligence." Consequently, the ability of a group to perform consistently (well) across tasks is defined as "collective intelligence" (CI) (Aggarwal et al. 2015; Engel et al. 2014; Woolley et al. 2010). Woolley et al. (2010) measured CI using an exploratory factor analysis in which a single general factor explained substantial variance in a group's performance outcomes across various tasks. Thus, we consider CI as an explanatory factor for a group's consistent performance across tasks. Such a CI factor is suitable for predicting how the group will perform.

Graf-Drasch et al. (2022) analyzed prior research on the CI factor and discovered that CI depends on the task structure. Task structure can be well-defined (having verifiably correct solutions) or ill-defined (lacking predetermined solutions) (Jonassen 1997; Schraw et al. 1995). Across well-defined tasks, groups exhibit CI. That is, groups that perform well on well-defined tasks tend to perform similarly on other well-defined tasks. Across ill-defined tasks, groups do not exhibit CI, and it is not possible to infer performance on one task from performance on other tasks (Graf-Drasch et al. 2022).

Going beyond the distinction based on the task structure (well-defined and ill-defined) might allow an even better understanding of group performance and its predictability. Coordination theory suggests that people perform two core activities when collaborating in groups regardless of the task structure (Malone and Crowston 1990, 1994). These two core activities are production and coordination. Production activities are performed individually. Hence, they are likely related to individual intelligence (Barlow and Dennis 2016a). In contrast, coordination is about managing the interdependencies of group members. Coordination is not, or at least to a lesser extent, related to individual intelligence. Usually, well-defined tasks tend to be high in production, and ill-defined tasks tend to be high in coordination (Graf-Drasch et al. 2022). Nonetheless, there are ill-defined tasks, such as brainstorming activities, that require little to no coordination because group performance is the sum of individual effort. In a similar vein, coordination activities also occur in well-defined tasks.

Based on coordination theory, Barlow and Dennis (2016a) argued that group members' individual intelligence is a predictor for group performance in high-production tasks, as group performance is largely the sum of individual work. In other words, the group members' intelligence is a factor that explains the performance (for high-production tasks) and, hence, is a predictor for performance and CI. Such a relationship between group members' individual intelligence and their CI has also been shown in the seminal study by Woolley et al. (2010) but not in all subsequent work on CI. For group tasks that require higher coordination among group members, the intelligence of individual group members only has a limited impact on the outcome of group performance (O'Brien and Owens 1969) and, hence, does not seem to be a predictor of CI (Barlow and Dennis 2016a).

This effect cannot be explained solely by the work of Graf-Drasch et al. (2022), as the task battery used by Barlow and Dennis (2016a) comprised two well-defined tasks and one ill-defined task. To the best of our knowledge, all studies finding a CI factor used task batteries with a major share of well-defined high-production tasks (c.f. "Appendix 1"). The case is different for ill-defined high-coordination tasks, as groups did not yet exhibit such an explanatory factor in prior research. The remaining two task types (well-defined high-coordination tasks and ill-defined high-production tasks) have rarely been investigated concerning intelligence and predictable performance. Therefore, we suggest that the predictability of task performance does not simply rely on the task structure (well-defined or ill-defined) but also on the extent of required coordination (production or coordination).

While the structure of a task cannot be altered—an ill-defined task will remain one—the core activities of the task can be influenced. An example of this is the work of Barlow and Dennis (2016a), who reduced the coordination in the group in favor of production activities by task-specific preparation to strengthen the impact of individual intelligence on group performance. Consequently, this work focuses on well-defined high-coordination tasks and how to facilitate group work so that performance is somewhat predictable from group characteristics, specifically from group members' individual intelligence, as an explanatory factor. Going beyond the task-specific preparation used by Barlow and Dennis (2016a), we suggest a scalable, task-type-specific (not task-specific) approach. Thus, we pose the following research question (RQ):

> **RQ** How to facilitate group work in well-defined high-coordination tasks so that group members' individual intelligence is an explanatory factor for group performance?

We propose automated facilitation that affects the core activities. Such automated facilitation can be implemented in a group support system (GSS) or other digital collaboration environments. A GSS is a tool that can help groups to structure coordination processes without changing the nature of the task (Barlow and Dennis 2016a). We conduct an online experiment with 64 virtual groups performing various tasks while communicating via text-based chat, a common knowledge work setting. Results from our online experiment indicate that automated facilitation is an approach leading to the emergence of an explanatory factor for group performance in well-defined high-coordination tasks, fostering predictability of group performance.

The paper is structured as follows: Sect. 2 provides the theoretical background on group performance and the influence of intelligence, task types, and structures, and facilitating collaboration in GSS and other digital collaboration environments. It also presents the hypotheses posed about facilitating group work, depending on the type of task. We outline our experiment design and procedures in Sect. 3. Section 4 presents the results of our quantitative analysis. After discussing our results and their implications, we give an outlook addressing limitations and future research in Sect. 5. In Sect. 6, we conclude with a summary.

2 Theoretical Background and Hypothesis Development

2.1 Group Performance and the Influence of Intelligence

Individual performance across tasks is strongly linked to individual intelligence (Deary 2000; Gottfredson 1997; Spearman 1904). However, in group work, we note a fundamental decoupling of performance and intelligence in most tasks (Barlow 2017; Barlow and Dennis 2016a, b; Day et al. 2004; Woolley et al. 2010). Prior studies on how individual intelligence affects task-specific group performance reveal that individual intelligence is more substantial in additive tasks than in compensatory tasks where group members consider multiple criteria (Devine and Philips 2001; Steiner 1972; Valacich et al. 2006). In additive tasks, group performance is essentially the sum of the individual performances of all group members. For group tasks that require higher coordination among group members, the intelligence of individual group members only has a limited impact on the outcome of group performance (O'Brien and Owens 1969), and other aspects seem to play a role (Dennis et al. 2022; Woolley et al. 2010). This limited impact on group outcomes suggests that these outcomes fundamentally differ from the outcomes of individual tasks (Barlow and Dennis 2016b; O'Brien and Owens 1969).

Woolley et al. (2010) took the basic idea that individual intelligence drives individual performance across tasks and applied this concept to groups, assuming that groups, like individuals, have a characteristic level of intelligence (Woolley et al. 2010). They conceptualized CI as a group's ability to perform well on various tasks. This definition led to a new stream in the research on CI (Barlow and Dennis 2016a; Gimpel 2015; Graf-Drasch et al. 2022; Kim et al. 2017). With CI, group performance tends to be consistent across tasks. Groups that perform well (or poorly) in one task also perform well (or poorly) in other tasks (Woolley et al. 2010).

Prior research has investigated the antecedents and boundary conditions of CI. The literature has suggested that CI is influenced, to some extent, by the general individual intelligence of the group's members (Barlow 2017; Barlow and Dennis 2016a; Bates and Gupta 2017; Woolley et al. 2010). Other individual-level factors, such as group members' social sensitivity, personality, and group collaboration factors, such as balanced speaking turns, may be relevant for the groups to exhibit CI (Dennis et al. 2022; Woolley et al. 2010). The medium for collaboration (face-to-face or computer-mediated) does not seem to have an influence (Engel et al. 2014, 2015; Meslec et al. 2016; Woolley et al. 2017). Other studies, however, have not found empirical evidence for CI in the vein of the Woolley et al. (2010) study. Summarizing this prior work, Graf-Drasch et al. (2022) recently pointed out that the task structure enables or inhibits groups exhibiting CI.

2.2 Task Types and Structure

Prior research has posited that a group's ability to perform a task well depends on the basic structure of the task and core activities (Graf-Drasch et al. 2022; Kitchner

1983; Newell and Simon 1972; Schraw et al. 1995). With regard to *structure*, a task is either well-defined or ill-defined. Well-defined tasks have evident solutions (Jonassen 1997; Schraw et al. 1995) that are verifiably correct, and there is a guaranteed procedure to achieve the solution(s) (Jonassen 1997; Schraw et al. 1995; Simon 1973). In contrast, ill-defined tasks lack predetermined or verifiably correct solutions or procedures (Jonassen 1997; Schraw et al. 1995). They require a unique solution different from other tasks' solutions (Newell and Simon 1972).

When people work in groups, they perform two *core activities*: production and coordination (Malone and Crowston 1990, 1994). Production is performed individually and, therefore, likely to be related to individual intelligence rather than interdependencies among group members (Barlow and Dennis 2016a). In contrast, coordination is about managing the interdependencies of group members. Coordination involves four different activities: communication, perception of common objects, group decision-making, and (pure) coordination (Malone and Crowston 1990). Group tasks typically require both production and coordination, although some tasks are particularly high in coordination, while others are primarily production-oriented (Kittur et al. 2009).

Although related, task structure and core activities are two concepts that lead to four task types. First, ill-defined tasks usually require considerable effort to understand the task (Simon 1973). Therefore, ill-defined tasks are often high in coordination. Moral reasoning is an example of such an ill-defined task that is high in coordination. Second, not all ill-defined tasks are high in coordination. There are ill-defined tasks that require substantial production, such as creativity tasks like brainstorming. Third, well-defined tasks typically are rather low in coordination. Copying a text is an example of a well-defined task where the correct solution is the sum of individual efforts. Fourth, well-defined tasks might also require a higher share in coordination activities (Graf-Drasch et al. 2022). An example is an estimation task where the group must correctly answer a numerical estimation question. Table 1 illustrates examples of the four different task types.

In summary, a task can be conceptualized as having two dimensions: the core activities and the task structure. The core activities can be either production or coordination. The task's structure relates to the task's solution or expected response. Well-defined tasks have verifiably correct answers, while ill-defined tasks have multiple solutions or no apparently correct solution. Both dimensions should be considered because they both influence the performance of groups. Therefore, we propose considering four task types: ill-defined high-production tasks, well-defined high-production tasks, ill-defined high-coordination tasks, and well-defined high-coordination tasks.

Extending on this differentiation, the task types also relate to the inherent complexity of performing the tasks. Snowden and Boone (2007) define four types of complexity. Presumably, complexity results from both the core activities and the solution to the task, affecting whether groups can exhibit a CI factor. For well-defined high-production tasks, it is reasonable to conclude that this is a simple context (in the terminology of Snowden and Boone 2007). For example, the right

Table 1 Four task types with examples

		Task structure	
		Ill-defined	*Well-defined*
Core activities	*High-production*	**Ill-defined high-production tasks** Example: Brainstorming. The group works out several correctly brainstormed items (e.g., words or numbers fulfilling given criteria) that can be added up (i.e., high-production) once duplicates have been removed. The task is ill-defined due to many possible solutions. Type of complexity: *complicated*	**Well-defined high-production tasks** Example: Copy text. The group must copy text into a shared document. This task has a clear process structure: typing a text. The result is equally clear: the text to be typed (i.e., well-defined). The task is additive in that the number of correct characters produced by each group member can be summed up (i.e., high-production). Type of complexity: *simple*
	High-coordi-nation	**Ill-defined high-coordination tasks** Example: Moral reasoning. The group evaluates arguments regarding two moral dilemmas following no specific pattern; instead, the procedure for solving the task is arbitrary (i.e., ill-defined). More people cannot provide more answers, as the participants' stringency in the evaluation of the arguments is rated. A lot of coordination is required to reach a joint decision (i.e., high-coordination). Type of complexity: *complex*	**Well-defined high-coordination tasks** Example: Estimation. The group must correctly answer a numerical estimation question. The correct answer can be reached by successive guessing (i.e., well-defined). Since it is an estimation task for which group members must discuss their opinions, their results cannot simply be added up (i.e., high-coordination). Type of complexity: *complicated*

answer is undisputed, and all parties share an understanding. In such simple tasks, prior research repeatedly found a CI factor, suggesting rather high consistency and, thus, predictability of performance across tasks. In contrast, ill-defined high-coordination tasks are arguably the most challenging task type and presumably a complex context, as it is not possible to easily discover correct answers (for completeness, chaotic contexts are not of relevance here). In complex contexts, a single action or decision can cause flux and affect succeeding decisions, exhibiting previously unknown interrelationships that make it difficult to track down correct answers. It is a hindrance for groups to exhibit a CI factor. Well-defined high-coordination tasks and ill-defined high-production tasks in between in terms of complexity that is, complicated. Not everyone recognizes the existence of a clear relationship between cause and effect (i.e., which answer is adequate for the task) in well-defined high-coordination tasks. Ill-defined high-production tasks may encompass multiple correct answers. In both cases, "known unknowns" (Snowden and Boone 2007) play a role that affects how to respond to the respective task. Whether groups can exhibit a CI factor in complicated contexts is unclear due to the lack of available research (cf. Sect. 1). In sum, this would suggest that, with the lower complexity of the task, groups tend to exhibit CI.

2.3 Facilitating Collaboration in Group Support Systems and Other Digital Collaboration Settings

Collaboration must be facilitated to effectively support and enhance the collaboration process (French 2013). Previous research has shown that the group collaboration process, encompassing how the group collaborates, holds significant importance (Riedl et al. 2021; Woolley et al. 2010). Consequently, it is important to support the interaction processes of the group (e.g., nudging the group toward a more effective group coordination strategy). Such facilitation can be seen as a set of functions or activities carried out before, during, and after a group interaction to help the group achieve its intended outcomes more easily (Bostrom et al. 1993). Facilitation is a dynamic process that involves managing relationships between people, tasks, and technology and structuring tasks to achieve intended outcomes. Facilitation can be implemented in a GSS or other digital collaboration environments.

Facilitating group work with GSS is a topic with a long history in information systems research (e.g., Chen et al. 2007; Galletta and Zhang 2014; Kilgour and Eden 2010; Newman and Dynamics 2021; Pervan 1998). Traditionally a GSS "aims to improve the process of group decision-making by removing common communication barriers, providing techniques for structuring decisions and systemically directing pattern, timing, and content of the discussion" (DeSanctis and Gallupe 1987, p. 589). As this common definition dates back to 1987, removing communication barriers, from a technical perspective, is of secondary importance today as social computing tools, including instant messengers, (group) chats, videoconferencing tools, and shared workspaces and documents, contribute significantly. However, the remaining aspects of facilitating group work (communication barriers from a social perspective, techniques for structuring decisions, and guidance in discussions) are still relevant and correspond to what facilitation means (Lim and Guo 2008; Nunamaker et al. 1991). An important factor for the success of a GSS is that it facilitates the aspect of collaboration (e.g., coordination) that is relevant for completing the task, respectively, the type of task (Zigurs and Buckland 1998). The management and use of GSS are closely related to facilitation.

A GSS eases access to information and understanding of the group processes, and it is important to recognize that the facilitator plays a key role in shaping and guiding the group process and the design and use of the GSS. For facilitation to be effective, facilitators must be able to monitor (a substantial part of) a group's communication and collaboration activities. Due to its complexity and variety, it used to be an activity carried out by humans only. However, human expert facilitators are scarce due to the need for technological proficiency and an understanding of group dynamics (Nunamaker et al. 1996). To resolve this conflict, digital collaboration environments, such as shared cloud storage or collaborative document editing, which are common in many contemporary work settings, offer the potential for digital facilitation and its special form, automated facilitation (Limayem et al. 1993).

Automated facilitation means "emulating a human facilitator in guiding groups" (Limayem et al. 1993, p. 98) and has been investigated recently from multiple perspectives, such as necessary capabilities (Bittner et al. 2021; Gimpel et al. 2023; Tavanapour and Bittner 2018) to prototypical development and evaluation (Gimpel et al. 2020; Kim et al. 2020; Przybilla et al. 2019; Winkler et al. 2019; Winkler and Roos 2019). Automated facilitation has the potential to alleviate the administrative burdens placed on human facilitators, resulting in cost reduction and potential improvements in group work efficiency and effectiveness at scale (Wong and Aiken 2003). While a human facilitator remains essential in certain situations, automated facilitation becomes increasingly important in virtual group work, especially when teams are geographically dispersed. In circumstances where deploying a human facilitator might not be feasible or disproportionate, automation offers the ability to facilitate at scale, addressing challenges traditional methods face in ensuring effective coordination across large and/or spread-out teams (Gu et al. 2021). The rationale advocating for automated facilitation is comparable to the intentions behind collaboration engineering, i.e., creating and implementing repeatable collaboration processes for collaborative tasks, leveraging facilitation techniques and technology (Kolfschoten et al. 2006).

Facilitation activities can typically be classified as content or process facilitation (Bittner et al. 2021; Bostrom et al. 1993; Dennis and Wixom 2002; Khalifa et al. 2002). *Process facilitation* focuses on the processes and relationships within the group, and a facilitator, in this sense, acts as a guide to manage, structure, and simplify these processes and manage interpersonal tensions (Chan et al. 2016; Dennis and Wixom 2002; Ito 2018). In the context of this research, this may be particularly necessary when groups behave sub-optimally in their coordination, either because they have developed unfavorable routines over time or because they have not yet established routines for collaboration and lack a strategy. Overcoming sub-optimal behavior requires adoption by the group members, and process facilitation is particularly important until the group has incorporated the new process (Dennis and Wixom 2002). If, for example, a group coordinates too much in a high-production task, they may lose valuable time. So, the group needs to become aware of which group activities are appropriate and to which extent for which task. Gupta et al. (2019) facilitated group work to foster CI through a chatbot, which sent static, predefined messages to address specific team processes. In particular, they nudged the group so that the most-skilled member contributed as much as possible to their production-heavy tasks. A second example is Barlow and Dennis (2016a), who designed a tool to reduce the coordination effort by the group to gain predictability through the impact of individual intelligence as the task becomes more production-heavy. In high-coordination tasks, several coordination issues may occur (Dennis 1996; Dennis et al. 2008; Thissen et al. 2007), all of which may inhibit the impact of individual intelligence on virtual group work (Barlow and Dennis 2016a). For coordination tasks, the challenge is to facilitate the coordination aspects, such as strengthening the relationship between group performance and individual intelligence (Barlow and Dennis 2016a).

Content facilitation involves actions that directly influence the content of the group's work (Chan et al. 2016; Clawson and Bostrom 1996), such as expressing one's opinion, creating awareness for connections with others, and strategies, such as fostering divergent thinking to avoid homogenous knowledge creation (Chan et al. 2016; Clawson and Bostrom 1996). The facilitator is rather an expert participant. As our approach is to provide task-generic support, this dimension is not in the scope of our automated facilitation. However, in combination, content and process facilitation are supposed to have a multiplying effect (Eden and Radford 1990). Our focus is to support groups applying favorable strategies when coordinating, depending on the type of task.

2.4 Hypotheses

In this research, we study the facilitation of group work so that performance becomes predictable by an explanatory factor. In our study, we focus on group members' individual intelligence as an explanatory factor rather than on CI for two reasons. First, CI is defined as "the general ability of the group to perform a wide variety of tasks" (Woolley et al. 2010, p. 687). The restricted domain of well-defined high-coordination tasks is not varied enough to do justice to this broad definition. Second, the automated facilitation partially shifts the groups' core activities from coordination toward production, with performance on production activities related to individual intelligence (Barlow and Dennis 2016a). Hence, relating group performance to individual intelligence is close to the theorized mechanism at work. As individual intelligence is an individual trait with rather high consistency over time, a relation of individual intelligence to group performance results in consistent and predictable group performance across tasks.

We outlined in Sects. 1 and 2.2 why it is important to consider the task type (i.e., the interplay of task structure and core activities). In the following discussion, we elaborate, for each of the four task types, whether an explanatory factor for consistent group performance should exist regarding the task's conceptualization as having two dimensions, resulting in four different task types. First, research agrees that for ill-defined high-coordination tasks, that is, (likely) complex contexts, no explanatory factor for consistent group performance exists (Barlow and Dennis 2016a; Graf-Drasch et al. 2022).

Second, the case is unclear for ill-defined high-production tasks, and we cannot make any reliable statements on such complicated contexts *ex-ante*. While the study of Barlow and Dennis (2016a) suggested that a CI factor is present and related to individual intelligence in high-production tasks, the meta-analysis by Graf-Drasch et al. (2022) examined 21 studies (including Barlow and Dennis 2016a) and suggested that groups do not exhibit CI for ill-defined tasks. The ill-defined tasks in the 21 studies differed in terms of their degree of coordination and production (refer to "Appendix 1" for details). Consequently, the findings of Graf-Drasch et al. (2022) have not allowed a clear statement regarding whether groups in general do not exhibit a CI factor in ill-defined tasks (whether high-production or high-coordination) or only do not exhibit a CI factor in ill-defined high-coordination tasks and do exhibit a CI factor ill-defined high-production tasks. Hence, we cannot derive from the literature whether groups exhibit a CI factor for ill-defined high-production tasks. Notably, the studies in which groups did not exhibit a CI factor involved a greater proportion of ill-defined high-production tasks than ill-defined high-coordination tasks. Thus, we hypothesize:

H1 For ill-defined high-production tasks, there is no substantial association between individual intelligence and group performance.[1]

Third, once a high-production task is not ill-defined but well-defined (i.e., simple context), the prior research has agreed that groups exhibit a CI factor (Barlow and Dennis 2016a; Graf-Drasch et al. 2022). According to Barlow and Dennis (2016a), this CI factor correlates with the individual intelligence of the group members. Although the following hypothesis is not expected to expose new insights, it strengthens the concept and the complete consideration of the four task types. As a consequence, we hypothesize:

H2 For well-defined high-production tasks, there is a positive relationship between individual intelligence and group performance.

Fourth, the case for well-defined high-coordination tasks is not clear. Following the argumentation above, the meta-analysis by Graf-Drasch et al. (2022) suggested that groups exhibit a CI factor for well-defined tasks. In this case, the work of Barlow and Dennis (2016a) was not contrary but allowed for limited conclusions. Their probands worked on three tasks (two well-defined and one ill-defined task, one low-coordination and two high-coordination tasks). Their treatment was an information system that reduced coordination issues between group members, making the respective task heavier in production. They achieved this by designing the tool specifically for each of their tasks, which required preparation for every task and was not generalizable to other tasks. Although their technical system focused on facilitating group work across multiple tasks, it was nonetheless task-specific. As a result, much effort was required to prepare each task. Barlow and Dennis (2016a) demonstrated that groups exhibit a CI factor through their treatment. The individual intelligence of groups explained the factor. Leveraging the individual intelligence of group members strengthened the impact that individual intelligence had on group performance and, thus, enabled groups to perform more consistently across tasks (Barlow and Dennis 2016a). While the structure of the task remained (two well-defined and one ill-defined task), the performed core activities were altered through the treatment. Core activities were significantly altered from coordination toward higher production. In the control group, this was not the case. Based on their treatment, it seems possible to address coordination within the group. As a consequence, a more production-heavy setting allows groups to exhibit a CI factor, which is explained by individual intelligence.

[1] The wording “no substantial association” was chosen with regard to our ability to empirically test H1 with experiment data. Achieving the statistical power to demonstrate the absence of any association would require a prohibitively large sample size. Thus, we put forward a cautionary hypothesis only, suggesting the absence of a substantial association. In the analysis of the experiment data, we tested for the absence of a medium-effect size.

With this study, we provide a generalizable approach that does not change the core activities for the task but still helps a group to coordinate successfully Our approach is different from the approach of Barlow and Dennis (2016a) in that it is not task-specific but applicable to a task type and needs no time-intense preparation for each task. Hence, the automated facilitation can be an easily adaptable solution within a GSS or other digital collaboration environments, as it is not only designed for a specific task but for different task types and, thus, suitable for numerous tasks. The facilitation focuses on offering advice that takes account of the dynamic nature of each group's collaboration. Thus, a collaboration system using this type of automated facilitation is 'generic' rather than task-specific since this promises consistent performance. As a consequence, we pose the following:

H3 For well-defined high-coordination tasks, the relationship between group performance and individual intelligence is positive and stronger in groups that use appropriate automated facilitation than in those that do not.

By addressing the effort of coordination, one can increase the positive impact of individual intelligence on the (consistency of) group performance (Barlow and Dennis 2016a). This unique approach, then, looks at the challenge of effortlessly assisting group work using technology. It does so from a new perspective by focusing on how to increase individual intelligence's positive impact on group performance by facilitating coordination activities.

3 Experiment Design and Procedures

We ran an online experiment with 64 virtual groups, each composed of three members. There were 31 treatment groups that received facilitation in the form of chat messages (c.f. "Appendix 2"). At the same time, 33 control groups did not receive facilitation, i.e., the advice was not active. Each group member performed an individual intelligence test, and each group performed the same four group tasks. We limit the research scope to virtual group work because communication and individual work and interventions by the automated facilitator are substantially easier to perform in a digital collaboration environment. These include collaborative text editing in a word-processing software with cloud storage and add-on communication via text chat. While it is theoretically possible in a non-virtual environment, for example, a group collaborating face-to-face with a voice assistant in a smart speaker, such as Amazon's Alexa in an Echo device, we consider such scenarios far-fetched. Therefore, we concentrate on digital collaboration, where most communication and individual work occur via digital media, particularly via text chat. Furthermore, online experiments with virtual groups are established in CI research since CI transcends media (Graf-Drasch et al. 2022). The University Ethics Committee approved the research (GfeW 2021).

3.1 Participants and Procedures

The online experiment's participants were 192 U.S. citizens (58% female, average age of 42 years) recruited via Amazon Mechanical Turk (MTurk), a crowd labor marketplace. In deciding on MTurk, we were well aware of the challenges that come along with it (Aguinis et al. 2021). Groups that work together over a long time are likely to develop routines, and adapting to new routines is potentially difficult. Consequently, we intended to investigate a baseline performance and put together groups that had not collaborated before and, thus, presumably, were rather open to the advice they were given. We assumed that, in principle, effects observable for a newly composed group would also be transferable to established groups when they experienced the benefits of automated facilitation. This should help the group to overcome sub-optimal behavior. However, established routine behavior and lacking novelty in the facilitation might weaken the effect. Here, further research on long-term usage is required. To ensure high-quality data, we restricted participation to workers who had already performed at least 1,000 other tasks on MTurk (MTurk wording: Number of HITs Approved) with a success rate above 96% (MTurk wording: HIT Approval Rate). Despite the considerable popularity of MTurk as a source of experimental data, some have questioned the sample quality. Such concerns notwithstanding, several researchers well-versed in testing sample quality have argued for the good and often superior quality of MTurk samples compared to regular internet panels and more traditional data collection methods (Buhrmester et al. 2016; Hauser and Schwarz 2016).

Before the online experiment, participants were randomly assigned to groups of three, and each group was selected randomly for either automated facilitation (treatment) or not (control). Hence, we had rather egalitarian groups where politics and power considerations did not play a major role, as participants did not know each other. At the start of the experiment, each group member performed an intelligence test. We used the 30-question WPT-Q version of the Wonderlic Personnel Test (Wonderlic 1992). It is a short yet valid version that has been applied in several prior studies (Hendy and Biderman 2019; Rakhmanov and Dane 2021; Tews et al. 2011). At the end of the online experiment, participants reported their demographics. The participants did not differ significantly between the treatment and the control group regarding their members' age, gender, and intelligence.

We used four group tasks from the McGrath (1984) circumplex to measure group performance reflecting a task related to real-world collaboration. Specifically, we chose the tasks to represent the abovementioned task types. In compliance with task characteristics and structures, we used *brainstorming* as the ill-defined high-production task, *group typing* as the well-defined high-production task, *estimation* as the well-defined high-coordination task and *moral reasoning* as the ill-defined high-coordination task (c.f. "Appendix 3"). Each task was timed as typical for group performance experiments. The entire group task battery was completed in under 1 h. While real-world tasks may stretch over a longer period, they can be broken down into smaller pieces. For example, a product and its components are broken down into epics, features, user stories, and tasks in software development. Real-world tasks may be executed in parallel, as one might have to wait for input. The rationale of the online experiment was to break down tasks into a small but still realistic unit that allowed measurement of performance in an adequate timeframe. The order of tasks varied between groups to avoid order effects. Each participant received a participation fee (6 USD) and a bonus based on group performance (0–1 USD for each group task) to incentivize performance.

3.2 Design of the Automated Facilitation

The automated facilitation used in the online experiment was a chatbot that provides advice to participants and is embedded in a shared online web application as the digital work environment for the experiment. The shared online web application was designed so that common, computer-based work activities required in the experiment could be carried out. Figure 1 depicts an exemplary screenshot. The participants could see a description of their task, information on the timing of the task, a chat, and a view of a shared document where they collaborated. The first two elements were only important for administering the online experiment and guiding the participants. The latter two were used daily by thousands of employees. (Examples in real-time online collaboration are Microsoft Office365 and Google Docs, and Zoom and Slack in communication). The shared online web application ensured that all group members worked on the same tasks simultaneously and could see the inputs of their fellow group members in real-time (similar to the procedure used by Barlow and Dennis 2016a, Gupta et al. 2019, or Engel et al. 2014). This aspect aimed to remove the barriers to shared communication, even if this was due to many comparable online services now of secondary importance. The participant could freely choose their display name in the chat and remain anonymous. However, their display name was shown with every message they sent.

The automated facilitation was performed by a chatbot advising the participants to support them in making decisions and structuring their discussions. Stasser and Titus (2003) conducted extensive work on hidden profile outlines in decision-making situations. They found that once information was not fully symmetric within a group, coordination and facilitation of information were important, as this can severely impact the outcome. Hence, the automated facilitation focused on providing advice considering the dynamics of each group (cf. Sect. 3.2.1). Although more factors (such as hierarchy, personality, or sympathy) might affect how individuals collaborate, this work primarily focused on facilitating group work by guiding groups in using appropriate group activities for a task type. Thus, the automated facilitator supported groups by giving scheduled advice on best-performing strategies for each task type (i.e., not for the individual task). Gupta et al. (2019) followed a similar strategy. Regardless of the type of task, they designed two static messages sent to the participants to foster favorable collaboration strategies. Their chat nudge led to the desired group strategy (the most-skilled members of the group contributed most to the task's completion). However, six were well-defined high-production tasks where such a strategy (where no coordination is needed) was appropriate. The remaining two tasks were ill-defined. Hence, their proposed solution was not applied to well-defined high-coordination tasks.

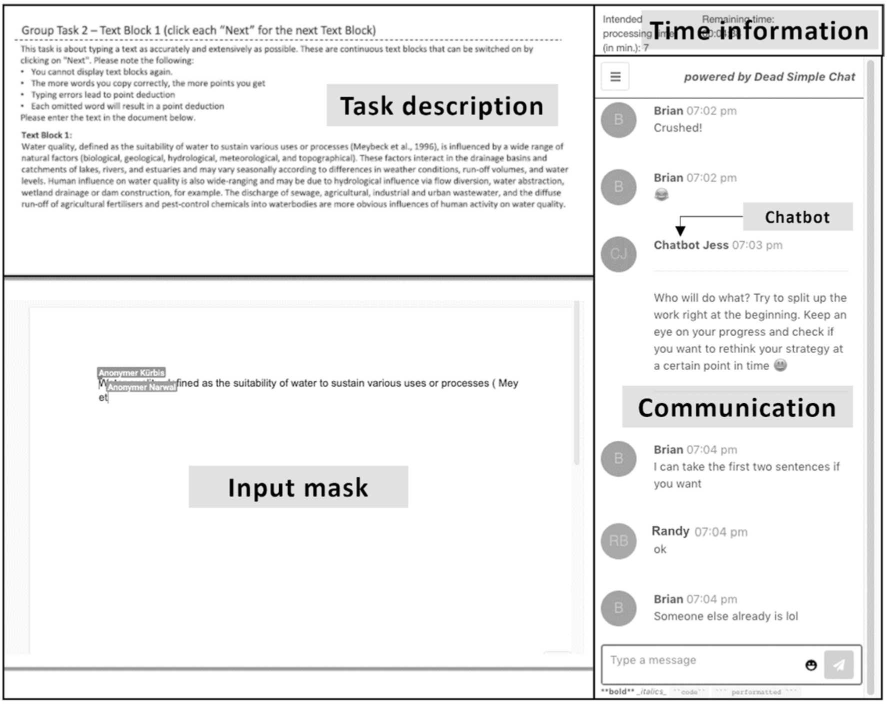

Fig. 1 Shared online web application displaying Group Task 2 (group typing task: well-defined high-production)

Therefore, our chatbot aimed to facilitate group work, particularly in well-defined high-coordination settings, by equipping a group's members with the means to coordinate more effectively, resulting in a higher share of production over coordination activities and, hence, a more predictable performance. Its rationale was comparable to what was intended in collaboration with engineering's ThinkLets (Kolfschoten et al. 2006). However, the variety of ThinkLets and corresponding preconditions is too detailed when it comes to generic strategies for well-defined high-coordination tasks. For comparison, we also considered the other three task types and developed corresponding advice that was given to the participants. In Sect. 2.4, we outline what we expected to observe.

3.2.1 Development of Supporting Advice

To integrate the advice into the working process of the virtual groups, the chatbot sent each piece of advice as a message in the group chat. We used the Wizard-of-Oz (WoZ) technique for the chatbot to avoid expense and development efforts (Landauer 1987; Wilson and Rosenberg 1988). In the WoZ technique, a human helper, the wizard, performs the functions as the program would execute them (Wilson and Rosenberg 1988). When demonstrated to be successful, implementing the solution corresponds to an automated facilitator (Limayem et al. 1993).

All advice offered by the automated facilitator was based on best-performing strategies. We developed these as follows. Prior research has experimented in laboratories with CI in groups without facilitation (Gimpel and Graf-Drasch 2023). The tasks in our online experiment were identical to tasks in this prior experiment but without the chatbot moderating the tasks. Prior research has dissected the work process and related the group activities of high-performing and low-performing groups. The data analysis process involved two authors examining the lab records and coding the data from the text data. Using authors as coders is a common research practice (e.g., Porter et al. 2019; Shewach et al. 2019). Work processes and group activities during task performance were categorized as production and coordination processes, resulting in a sequence of group activities (Blaß et al. 2023). This prior work investigated the sequence of group activities (cf. Malone and Crowston 1990, 1994) in the respective experiment. Based on this data, it developed generic strategies for group collaboration for ill-defined high-production and well-defined high-coordination tasks (Blaß et al. 2023). As this prior work only has focused on two of the four task types, we analyzed sequences of group activities in low- and high-performing groups the remaining task types (i.e., ill-defined high-coordination and well-defined high-production tasks) and identified recurring patterns, too.

Building on this work and orienting at automated decision guidance (Limayem and DeSanctis 1993), we formulate our advice per task type. For example, in an ill-defined high-production task (e.g., brainstorming), coordination between group members (e.g., group decision-making, cf. Malone and Crowston 1990, 1994) hinders good performance. This is because much time is lost in decision-making, which is irrelevant to solving the task. It is more purposeful when the group carries out more production activities, with each member working on the solution. In that specific case, backward guidance (as defined by Limayem and DeSanctis (1993) is appropriate to inform the group members that they are exposing unfavorable behavior for being successful with the task. The identified patterns of collaboration were associated with high or low performance (e.g., the occurrence of certain activities that should be avoided or a minimum proportion of a certain activity). For the well-defined high-coordination task type, the best-performing strategies were long phases of task-related communication and intense production activities (Blaß et al. 2023). Less production- and task-related communication and making group decisions over a longer period of time were the low-performing strategies. Consequently, we formulated the advice as follows: "Take time to think about each upcoming question by yourselves. Discuss it and then decide, but be sure you have your own thoughts first!" If the group carried out less production during the group work, the advice was, "Everything alright? Try to primarily focus on the task, think by yourselves about respective solutions, and then agree quickly on a common solution." Therefore, the advice was tailored to guide the group process so that the group performed the right group activities (or avoided inappropriate ones) to achieve consistent group performance.

The advice was set up as preventive guidance given at the beginning of the task and intended to hint at the appropriate group activities for that type of task. Alternatively, the advice was formulated as backward guidance when the group deviated from desired group activities, either as they performed activities that did not fit the type of task (i.e., a trigger, such as too much time on decision-making is fired) or as they did not perform favorable activities extensively enough (i.e., production in a high-production task). This process ought to reduce hidden profiles in high-coordination tasks. The group received forward guidance if no such trigger occurred, and the group should continue as they were. Drawing on Snowden and Boone (2007), we aimed to contribute in complicated contexts. The automated facilitator sensed the group activities, analyzed the fit to the given task, and responded if necessary. The following paragraph describes the design of the response in terms of language style and interaction mode.

3.2.2 Formulation of Supporting Advice

Advice on how to perform group tasks is at the core of automated facilitation (c.f. "Appendix 2"). A welcome message appeared in the group chat at the beginning of the first group task. This message comprised initial information on the chatbot's purpose.

For each task type (i.e., well-defined high-production, ill-defined high-production, well-defined high-coordination, ill-defined high-coordination), we formulated advice based on identified recurring patterns of best-performing and avoiding strategies described in Sect. 3.2.1. The appearance of subsequent advice depended on each group's working process, which was bound to conditions and timings. We tracked each group's task performance activities and compared them to the "best practice" durations of activities set by well-performing groups in a prior experiment. "Appendix 2" lists the thresholds that triggered the automated facilitator to give advice. "Appendix 4" describes the monitoring and the resulting triggers.

We formulated each piece of advice following established standards of human-chatbot interaction (Chaves and Gerosa 2019; Diederich et al. 2020). These standards suggest adding social characteristics to chatbot messages, like a name or including emojis and motivational phrases like "Good luck" or "Go for it" to make the chatbot more accepted by users. In addition, we integrated questions into chatbot messages because advice in the form of questions elicits favorable reasoning and elaboration behavior (Ito et al. 2022).

These automated facilitation messages were specific to the task type but not to the individual task, making them substantially more generic than the approach of Barlow and Dennis (2016a). We do not claim that these automated facilitation messages are optimal. Instead, we developed them to be effective, and if they were, they were sufficient to answer our research question.

4 Analysis and Results

To test hypotheses H1 to H3, we ran separate ordinary least squares regression models. We normalized the performance scores across tasks as the dependent variable and used each group's average intelligence as an independent variable. The activation of the chatbot served as a moderator, coded as a binary variable (0 for the control group, 1 for the treatment group having the automated facilitator). We used a

0.10 significance level deemed appropriate given the sample size of 31 treatment and 33 control groups. Table 2 summarizes the results.

Regarding ill-defined high-production tasks, our regression results (Model 1 in Table 2) indicate no significant relationship between average individual intelligence and group performance, whether the control or treatment groups used the automated facilitator. A post hoc power analysis suggests that, given our sample size and significance level, our analysis has a power of 0.92 to detect a medium effect (f^2 0.15 or larger). The beta error probability is 0.08—smaller than our chosen significance level. Hence, given the power of our analysis compared to the significance level, the absence of finding a significant effect of individual intelligence on group task performance can be interpreted as evidence of the absence of a medium or strong effect.[2] H1 hypothesized no "substantial association" exists between individual intelligence and group task performance. Hence, the results support H1.

Regarding well-defined high-production tasks (Model 2 in Table 2), our results indicate a significant relationship between average individual intelligence and group performance, whether in the control or treatment groups. The data supports H2.

Moving on to well-defined high-coordination tasks (Model 3 in Table 2), average individual intelligence does not significantly impact the control groups' performance. But—as shown by the interaction term—average individual intelligence has a notable positive impact on the performance of the treatment groups. The data supports H3.

For completeness, we also tested the ill-defined high-coordination task type, where we could uniformly conclude from the literature that no significant relationship exists between average individual intelligence and group performance. Our results support this, as no significant relationship occurred between average individual intelligence and group performance in the ill-defined high-coordination task (Model 0 in Table 2).

5 Discussion

Existing research has analyzed individual intelligence and CI as factors related to consistent performance across tasks (Barlow and Dennis 2016a; Graf-Drasch et al. 2022; Woolley et al. 2010) and has demonstrated that this depends on the task type (Barlow and Dennis 2016a, b; Fuller and Dennis 2009; Graf-Drasch et al. 2022). The task type comprises the task's fundamental structure (i.e., ill-defined or well-defined) and the core activities required to solve the task (i.e., coordination activity or production). Groups do not exhibit a CI factor for ill-defined tasks, regardless of whether a task is high-production or high-coordination (Graf-Drasch et al. 2022). For well-defined high-production tasks, groups exhibit a CI factor related to the group members' average individual intelligence.

[2] Identifying a small effect (f^2) would require a sample size of at least 430 three-person groups (1,290 individuals). We were not able to run the experiment with such a large sample, especially due to time requirements for the Wizard-of-Oz technique.

Table 2 Results of regression models for task performance as dependent variable (parameter estimates, p-values, and significance levels; n = 64 for all models)

	(0) Ill-defined high-coordination	(1) Ill-defined high-production	(2) Well-defined high-production	(3) Well-defined high-coordination
Related hypothesis	None	H1	H2	H3
Intercept	22.05 (0.001 **)	22.88 (< 0.001***)	745.90 (0.002**)	10.45 (< 0.001***)
Average Intelligence	− 1.22 (0.129)	0.81 (0.258)	58.98 (0.045*)	0.01 (0.973)
Treatment	0.31 (0.979)	− 9.65 (0.367)	− 104.98 (0.807)	− 8.04 (0.030*)
Average Intelligence*Treatment	− 0.26 (0.848)	0.33 (0.792)	− 11.90 (0.812)	0.73 (0.090 +)
R^2	0.08	0.12	0.11	0.12
Adjusted R^2	0.03	0.07	0.07	0.08

$^{+}p < 0.1$; $^{*}p < 0.05$; $^{**}p < 0.01$

We focused on facilitating group work on well-defined high-coordination tasks so that group members' individual intelligence became an explanatory factor for group performance. For comparison, we also considered the other task types. Well-defined high-coordination tasks are relevant as they comprise many important tasks, for example, forecasting projected profits for a company or estimating the expected increase in global warming (Stern et al. 2017). Having an explanatory factor somewhat predictive of group performance is important because it allows the assembly of consistently performing groups, an essential organizational activity that decreases the risk of poor performance.

Prior research has suggested that distinguishing tasks as well-defined or ill-defined matters to better understand group performance and its predictability (Graf-Drasch et al. 2022). Based on coordination theory, our first finding points out that the core activities required for a task—production or coordination—are complementary, relevant boundary conditions for the predictability of group task performance from intelligence. We find empirical support for this result. While we observe a significant influence of individual intelligence on group performance for well-defined high-production tasks, no such factor exhibits for well-defined high-coordination tasks. This result can be reconciled with previous research if most prior studies used a set of mostly well-defined high-production tasks when groups exhibited CI (c.f. "Appendix 1"). This raises the question of whether CI also occurs in a set of well-defined high-coordination tasks. Our results indicate this is probably not the case when their group work is not facilitated. Hence, task structure and core activities are relevant in the relationship between intelligence and group performance.

Second, well-defined high-coordination tasks can be facilitated so that a group is able to draw from its intelligence, allowing for the prediction of its performance. When groups are automatically facilitated (with a chatbot) for well-defined high-coordination tasks, there is a significant difference in how the average intelligence of group members relates to group performance. Groups without access to the chatbot do not exhibit such a relationship. Accordingly, when groups use the right tool to facilitate coordination, their intelligence affects performance outcomes and makes it possible to anticipate those outcomes. Barlow and Dennis (2016a) also addressed coordination issues by pre-structuring high-coordination tasks like production-focused tasks. The fundamental structure of the task was changed. Instead, we addressed the coordination effort during the task performance by offering advice to facilitate the group's working process. These pieces of advice were based on the best-performing strategies, formulated in general terms as they were specific to the task type but not to the individual task, making them more general than the approach of Barlow and Dennis (2016a) and better suited, particularly, to well-defined high-coordination tasks than the static advice given by Gupta et al. (2019).

5.1 Contributions and Implications for Research

First, our findings elaborate boundary conditions for the link between a group's performance and the individual intelligence of its members, which Table 3 highlights. Distinguishing a task beyond its structure (i.e., well- or ill-defined) allows for a better understanding of in which conditions groups exhibit CI. Our results foster the idea that four task types (well-defined high-production, well-defined high-coordination, ill-defined high-production, and ill-defined high-coordination) should be distinguished. Future studies should either point out when treatments and analyses are geared toward selected task types or ensure a balance between all four task types.

Second, the intelligence of individual group members determines performance in well-defined tasks. In production settings (as core activity), this is inherently given. In more complex high-coordination tasks, this is given when an appropriate corresponding facilitation is available. Hence, the second contribution of this study is that facilitating group work in a way that performance in well-defined high-coordination tasks relates to individual intelligence is possible in a dynamic manner that is not task-specific. This suggests that the facilitation of consistent group performance is possible across well-defined tasks (through automated facilitation independent of the core activities) and goes beyond simple tasks but also applies to complicated tasks. This contribution strengthens the theoretical conclusion of Barlow and Dennis (2016a) that when groups use the right technical solutions to coordinate their performance, their members' individual intelligence matters. While Barlow and Dennis (2016a) have used tools to reduce coordination and increase production efforts, we aimed to reduce coordination implicitly by guiding the group work.

Table 3 Relationship between a group's performance and its members' individual intelligence when being automatically facilitated (relationship for unfacilitated collaboration in parentheses)

Task type	Ill-defined	Well-defined
High-production	No (no)	Yes (yes)
High-coordination	No (no)	**Yes (no)**

Bold font refers to difference between facilitated and unfacilitated state

The process facilitation of a group for group activities alleviates a boundary condition for the effect of intelligence on group performance. Future research could refine the chatbot-based advice, incorporating the work on ThinkLets (Kolfschoten et al. 2006) and hidden profiles (Stasser and Titus 2003). In addition, we recommend further investigating ways to facilitate collaboration (how the group works together) in complex scenarios, as they require a high degree of coordination while dealing with multiple cognitive processes (Helquist et al. 2011).

The third contribution of this study concerns the design of automated facilitators. We provide such a facilitator for coordination issues in groups via chat-based communication. We offer an exemplary design in the form of messages related to specific triggers to facilitate coordination within the group, which leads to a positive relationship between individual intelligence and group performance in well-defined high-coordination tasks. The design includes the chatbot and, at its core, strategic advice to the group regarding their collaboration mode. Our automated facilitator in the context of a digital collaboration environment differs from existing GSS in that it is not task-specific, requiring preparation effort (Barlow and Dennis 2016a), but is applicable to a task type. It is not static and geared toward well-defined high-production tasks (Gupta et al. 2019) but is dynamic and suited for well-defined high-coordination settings reacting to the current group collaboration. Looking into the future in a next step, the WoZ technique could be implemented in combination with a state-of-the-art large language model that is able to sense group activities and generate messages that take into account the previous conversation while delivering the main message this research proposes, given the group activities. As a result, the automated facilitator could understand the group's activities (production and different coordination activities), based on the group's chat messages, with no need for a human facilitator. Looking further into the future, the chatbot could be an avatar joining a group's video call and giving the same advice.

5.2 Implications for Practice

This study focused on facilitating the coordination of group work by employing a chatbot to leverage the individual intelligence of group members and promote consistent group performance. Our findings reveal two practical implications. First, the individual intelligence of group members can have a greater impact on well-defined high-coordination tasks (e.g., decision-making) when appropriate facilitation is available. Given that functionality, such as our chatbot, is developed into a product, the setting could be replicated in practice (with adjusting timing and potentially framing of the triggers).

Substantial developments of individual assistance systems and group support are possible. Recent progress in the realm of natural language processing—especially based on large language models—allows for an ever more powerful integration of assistance systems in digital work environments (e.g., GitHub Copilot (Moradi Dakhel et al. 2023), Grammarly (Yuan et al. 2022). While individual assistance is easier to implement in information technology systems than group assistance, technical advances based on large language models are also likely to enter digital communication and collaboration environments, such as Microsoft Teams (Daderko 2023; Stallbaumer 2023). We envision that a "copilot" supporting communication and collaboration could support individual group members by note-taking and summarizing (Daderko 2023) and could also act as an automated facilitator. When groups collaborate via chat or in a phone or video call (for example, for a meeting or a specific topic), they could choose the type of task, or the type of task becomes evident from the context or content of the meeting. An automated facilitator could listen for triggers and support the group by offering advice based on best-performing strategies.

Obviously, our research did not explore the full richness of this scenario. We have not implemented comprehensive automated facilitation. Nevertheless, despite only being examples, as outlined in "Appendix 3", the tasks in our research were chosen to apply to a wide range of real-world tasks. Well-defined high-coordination tasks that were the focus of this study are relevant in many tasks, such as planning, resource management, resource allocation, and project management. Groups with more intelligent members are likely to perform various well-defined high-coordination tasks better than groups with less intelligent members. Hence, our research implies that it is worth exploring the implementation of automated facilitation of group work in digital work environments. However, it should not be neglected that collaboration, in reality, is probably a multi-faceted concept where social and political aspects play a role, groups might have a history of collaboration and established routines, and tasks are typically longer than in our experiment.

Second, when the relationship between group performance and individual intelligence is enhanced, group performance across tasks can become more consistent (Barlow and Dennis 2016a; Bates and Gupta 2017). Consistent performance has a significant impact on organizational group functioning. A group that performs well in tasks and can be depended on to do so in the future reduces risks. The manager can rest assured that the group is reliable in critical situations since they perform consistently well across tasks. There should be widespread interest in the effective organization of group work to enhance consistent performance. Managers seeking to build a reliable and consistently performing group for group work in an organization should apply automated facilitation that supports consistency once such functionality is available.

5.3 Limitations and Outlook

This study has limitations. One limitation may be that the online experiment was not conducted in highly controlled laboratory conditions or in the field. Instead, it was an online experiment during which participants may have been distracted and had no history of collaborating in the same group, which means they did not have established routines but also were able to act on par as there were, for example, no hierarchical differences. We included guidelines in our online experiment instructions to minimize distractions, such as "Set the full-screen mode for your browser" and "Do not start or quit any other programs." This allowed us to replicate the usual conditions of virtual group work, during which there could be distractions (Galluch et al. 2015; Rissler 2017). Online experiments have been established in CI research and conducted in multiple previous studies (Gupta et al. 2019).

A related aspect is that our research focused on text-only collaboration, as our main contribution was a chatbot aiming to facilitate the groups' working process (particularly for well-defined high-coordination tasks). It could be applied to a wider scope whenever a group works together via chat—as is quite common in the workday of a knowledge worker using communication tools. Whether a group could meet in the office was not important to our research as long as they could collaborate online. In larger companies, employees chat when they are in the office.

We chose an appropriate research design that matched our scope and focus, as described in Sect. 3.1, and followed previous research in this area. Nonetheless, we believe that further studies with long-term data may provide nuanced, valuable contributions that offset the following three limitations of our study. First, our study looked at groups working together for the first time and only once. Groups collaborating over a more extended period and solving similar tasks together may develop "implicit coordination skills" and task familiarity (Rico et al. 2008). In times of increasing agility in organizations, new groups often form. Individuals must adapt to working together on group projects. Future research should examine groups with a track record of collaborating on new and familiar tasks. Second, the online experiment's short duration, with four tasks in 1 h, limits the ability to assess longitudinal effects. Factors that impact group performance beyond the mere execution might be manifold. Group dynamics may evolve over time, influenced by various factors, such as trust-building and the development of shared norms and routines. Socio-political considerations, such as power, political dynamics, and personalities, can significantly affect collaboration and, hence, group performance. Power imbalances or conflicting personalities might hinder effective group work, as individuals might not contribute as they would in a neutral setting, eventually leading to sub-optimal performance. Hence, future research could aim to examine socio-political factors and their effects to further understand the predictability of performance. Third, it is essential to consider the temporal dynamics of adopting automated facilitation and acknowledge that the possible novelty-driven benefits (i.e., the emergence of an explanatory factor for group performance in well-defined high-coordination tasks, fostering the predictability of group performance) may diminish over time. Future research is necessary to examine the long-term effects of automated facilitation to determine whether the benefits experienced can be sustained through repeated and explicit encouragement to respond to the automated facilitator.

6 Conclusion

Our findings provide extended boundary conditions for the link between a group's performance and the individual intelligence of its members. We observed this link in well-defined high-production tasks and—when using a technical solution—well-defined high-coordination tasks, but not in ill-defined tasks. We provide an exemplary technical solution in the form of a chatbot dedicated to providing targeted, case-specific advice. The technical solution facilitates coordination issues in group work to establish the relationship between individual intelligence and group performance and achieve higher consistency, given individual intelligence as an explanatory factor.

Appendix 1

See Table 4.

Table 4 CI studies analyzed by Graf-Drasch et al. (2022, Table 2), extended by an analysis of core activities (production or coordination)—highlighted in bold

Research Paper	Sample size (number of groups)	Group size (range)	Group size (midpoint of range)	Number of well-, ill-defined tasks	Percentage of well-defined tasks (%)	Number of high-production, high-coordination, balanced production, and coordination tasks	Percentage of well-defined high-production tasks from well-defined tasks (%)	Exhibition of CI
Woolley et al. (2010)	40	3.0	3.0	3, 2	60	**3, 1, 1**	**60**	Yes
	107	2–5	3.5	6, 4	60	**6, 2, 2**	**60**	Yes
Engel et al. (2014)	32	4.0	4.0	6, 2	75	**6, 1, 1**	**95**	Yes
	36	4.0	4.0	6, 2	75	**6, 1, 1**	**95**	Yes
Engel et al. (2015)	116	2–5	3.5	5, 2	71	**5, 1, 1**	**95**	Yes
	25	4.0	4.0	5, 1	83	**5, 0, 1**	**95**	Yes
Barlow and Dennis (2016b)	86	3–5	4.0	0, 3	0	**1, 1, 1**	**0**	No
Barlow and Dennis (2016a)	64	3–5	4.0	2, 0	100	**0, 2, 0**	**0**	Yes
	65	3–5	4.0	0, 2	0	**2, 0, 0**	**0**	No
Meslec et al. (2016)	30	3–6	4.5	2, 1	67	**2, 0, 1**	**67**	Yes
Bates and Gupta (2017)	26	2–4	3.0	1, 2	33	**2, 0, 1**	**100**	No
	40	3.0	3.0	2, 3	40	**5, 0, 0**	**100**	No
Bates and Gupta (2017)	40	3.0	3.0	2, 3	40	**5, 0, 0**	**100**	No
Kim et al. (2017)	248	5.0	5.0	5, 1	83	**5, 0, 1**	**95**	Yes
Gimpel and Graf-Drasch (2020)	50	3.0	3.0	4, 4	50	**3, 2, 3**	**60**	Yes
	50	3.0	3.0	4, 4	50	**3, 2, 3**	**60**	Yes
	50	3.0	3.0	4, 4	50	**3, 2, 3**	**60**	No
	50	3.0	3.0	4, 4	50	**3, 2, 3**	**60**	No
Gupta et al. (2019)	136	3–4	3.5	6, 2	75	**6, 1, 1**	**96**	Yes
Ostrowski et al. (2019)	99	3.0	3.0	4, 0	100	**4, 0, 0**	**100**	Yes
Rowe (2019)	29	2–5	3.5	2, 3	40	**4, 0, 1**	**60**	No

Some research papers contain samples from multiple experiments. As experiment designs and results may differ between samples, we report them separately

Appendix 2

See Table 5.

Table 5 The pieces of advice provided by the automated facilitator

Task type	Advice threshold	Advice
Welcome message	At the beginning of the first group task	Hi there! I am chatbot Jess, and I will moderate your session. I provide advice to improve your performance. Please follow it as best as you can. Please note, I only give recommendations but do not answer any questions. Good luck:)
Well-defined high-production	At the beginning of this task	Who will do what? Try to split up the work right at the beginning. Keep an eye on your progress and check if you want to rethink your strategy at a certain point in time:)
	Perception of common objects > 15 s	If you have trouble handling your shared workspace, help each other
	Coordination ≤ 20 s	Are you happy with your progress and the way you work together? Discuss quickly if you want to change something
	Production ≤ 150 s	Try to get the job done. Coordinate briefly, then split up and work for yourselves. You will get it; I believe in you!
ill-defined high-coordination	At the beginning of this task	Take enough time to get an overview of the task and the worksheet. Get an understanding of what is expected and try to switch perspectives if necessary
	Task communication < 30 s	I think the best way to approach this task is to share your thoughts for each question and discuss related content. Then quickly decide. Let's try!
	Group decision making < 20 s	So far, so good, but I feel you're talking a lot about the reasoning for your answer to choose. You should also focus on the specific decision to make. Long story short: make decisions!
	Group decision making > 90 s	Hey guys! You're talking a lot about which answer to choose. How about focusing a little more on sharing your perspectives on the topic in order to decide. Explain your points of view!
Ill-defined high-production	At the beginning of this task	Take time to identify exactly what to do. Create a common understanding of the task. Talk briefly abou t the task and how you could organize your groupwork
	Group decision making > 0 s	Focus on the task and what you have to do. I feel as if you are a little off track
	Task communication < 10 s	Take a few seconds to share your thoughts about the task with each other. Give each other tips, share ideas. Even if it might be obvious, time is worth it. Maybe you will give an important hint
	Task communication > 20 s	Try to primarily focus on the task and think by yourselves about respective solutions. Coordinate briefly, then split up, and think and work for yourselves

Table 5 (continued)

Task type	Advice threshold	Advice
Well-defined high-coordination	At the beginning of this task	Take time to think about each upcoming question by yourselves. Discuss it and then decide but be sure you make your own thoughts first!
	Production < 20 s	Everything alright? Try to primarily focus on the task, think by yourselves about respective solutions, and then agree quickly on a common solution
	Group decision making > 50 s	Take a step back. Make your own thoughts first. Then be careful to make decisions quickly. Often that means not discussing things in too much detail. Go for it!
	Task communication < 30 s	Discuss about the topic of the question and make suggestions. Maybe you get another point of view, and this will be helpful:)
Midpoint message	if no other advice was triggered	How is your progress? Take time to reevaluate your status briefly to move forward. Well done so far!

Appendix 3: A Detailed Description of the Four Group Tasks Used Within the Online Experiment

Brainstorming—ill-defined high-production task (applied by, e.g., (Engel et al. 2014; Woolley et al. 2010): the group had 6 min to brainstorm and write down as many words as possible, each spelled with the first letter "L" and the penultimate letter "E." Every non-redundant correct word earned them one point. The task was not structured, so every member was told to write down eligible words as they occurred to them. Accordingly, there are many possible solutions (i.e., ill-defined). The task is additive since the number of correct words contributed by group members can be added up (i.e., high-production) once duplicates have been removed.

Brainstorming is a common and widely used technique in the working world to generate creative ideas and solutions. It allows individuals or teams to explore various possibilities, think outside the box, and come up with innovative solutions to complex problems. In the real working world, brainstorming is often employed during strategic planning, product development, problem-solving sessions, and team collaboration, where diverse perspectives and creative thinking are valued.

Group typing—well-defined high-production task (applied by, e.g., (Engel et al. 2014; Woolley et al. 2010): Participants had 7 min to transcribe a text by typing it into a shared document. The text was a scientific article split into multiple text blocks. One of these text blocks was always on display for the task's duration. Points were awarded according to the number of correctly reproduced characters. A score of 0–10,758 could be achieved per group. This task had a clear process structure: typing a text. The result is just as clear: the text to be typed (i.e., well-defined). The task is additive since the number of correct characters produced by individual group members can be added together (i.e., high-production).

While copying text may seem like a mundane task, it is important in the real working world (with some abstraction), especially for data entry, transcribing information, transferring content from one source to another, or programming. Specifically, many administrative and clerical tasks require accurate and efficient typing skills for tasks such as document preparation, report writing, data entry, and correspondence. Accurate and fast typing is critical to maintaining productivity and ensuring effective communication in a variety of professional settings. In somewhat more general terms, it also refers to the coordination of clearly defined tasks. Examples include writing program code for clearly defined interfaces, writing reports, or processing applications (e.g., credit checks or insurance claims).

Estimation—well-defined high-coordination task (applied by, e.g., (Engel et al. 2015; Woolley et al. 2010): Participants had 10 min to answer eight numerical estimation questions, each with four optional answers. Every correct answer earned three points, while the two answers closest to the correct response were each awarded one point. The answer furthest from the correct reply on the ordinal scale of four options did not earn any points. A score in the range of 0–24 could be achieved. Due to the provision of optional answers, the task outcome is clear: a selected answer possibility. The correct answer can be reached by successive guessing (i.e., well-defined). Since it is an estimation task for which group members had to discuss their opinions, their results cannot be added together (i.e., high-coordination).

Estimation is a valuable skill in the working world, as it allows individuals to make informed judgments and approximate quantities, costs, timelines, or resource requirements. Professionals frequently encounter situations where they need to estimate project budgets, timelines, resource allocations, or market demand. Accurate estimation supports effective planning, decision-making, and resource management, enabling organizations to set realistic goals, allocate resources efficiently, and meet project deadlines.

Moral reasoning—ill-defined high-coordination task (applied by, e.g., Bronikowska et al. 2019; Lind 2000): The Moral Reasoning Competence Test (Lind and Hartmann 1985) describes two moral dilemmas in which various arguments for or against behavioral responses must be evaluated. Participants had 20 min to solve the task. The score was assigned in conformity with the "C-value," according to (Lind and Hartmann 1985). The C-value is calculated as the portion of the variance in the participants' ratings attributed to the participants' concern about the moral quality of the arguments rather than, for instance, their agreement on opinions or other factors and combinations thereof (Lind 2000). A score in the range of 0–100 is achievable per group. Evaluating the various arguments does not follow a specific pattern; instead, the procedure for solving the task is arbitrary (i.e., ill-defined). More people cannot provide more answers, but considerable coordination is required to reach a joint decision (i.e., high-coordination).

Moral reasoning is relevant in the real working world, particularly in (ethical) decision-making and navigating complex (ethical) dilemmas. Professionals encounter situations where they need to make choices that align with ethical standards, uphold integrity, and consider the potential impact on stakeholders. Ethical reasoning helps individuals analyze implications, evaluate alternatives, and make principled decisions, which are vital in fields such as healthcare, law, business, and research, where ethical considerations are paramount.

Since the scores of the four tasks have very different ranges, task scores are normalized to a scale of 0–1 to ensure comparability.

Appendix 4: A Detailed Description of the Monitoring and the Resulting Triggers Belonging to the Automated Facilitator

Like Gupta et al. (2019), we found that the monitoring intervals of the automated facilitator to intervene must not be too close or too wide apart. While examining the results from previous studies (Blaß et al. 2023; Gimpel and Graf-Drasch 2023) to extract the strategies as described in Sect. 3.2.1, we found that 3–4 min are suitable for the automated facilitator to monitor the thresholds. On the one hand, the groups are not overloaded with advice. On the other hand, they get enough guidance to affect their collaboration process. With shorter monitoring intervals, activity phases are not separated, and if the intervals are too large, the advice may come too late and has no effect. Based on the monitoring, the automated facilitator sends a piece of advice if a group has strayed off the right path for too long. For example, if a group decision-making activity occurs in the ill-defined high-production task, the corresponding advice appears. If no advice was triggered, a midpoint message was sent instead. In such cases, the group activities were close to a best-performing strategy, so concrete advice was not required. The group was then motivated to continue using its successful strategy, albeit after a short pause for reflection.

Acknowledgements The research has been partly funded by the German Research Foundation (DFG, Grant Number 343128888) and is thus part of a comprehensive research endeavor.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aggarwal I, Woolley AW, Chabris CF, Malone TW (2015) Cognitive diversity, collective intelligence, and learning in teams. Collect Intell 1:3

Aguinis H, Villamor I, Ramani RS (2021) MTurk research: review and recommendations. J Manag 47:823–837. https://doi.org/10.1177/0149206320969787

Albert J (2004) Streakiness in team performance. Chance 17:37–43. https://doi.org/10.1080/09332480.2004.10554913

Barlow JB, Dennis AR (2016a) Group coordination structuring: a process to facilitate collective intelligence in virtual group work. In: Collective intelligence conference 2016, Boston, USA

Barlow JB, Dennis AR (2016b) Not as smart as we think: a study of collective intelligence in virtual groups. J Manag Inf Syst 33:684–712

Barlow JB (2017) Collective intelligence and its relationship to collective individual intelligence. In: Collective intelligence conference 2017, Boston, USA

Bates TC, Gupta S (2017) Smart groups of smart people: evidence for IQ as the origin of collective intelligence in the performance of human groups. Intelligence 60:46–56. https://doi.org/10.1016/j.intell.2016.11.004

Bittner E, Mirbabaie M, Morana S (2021) Digital facilitation assistance for collaborative, creative design processes: Hawaii international conference on system sciences 2021. In: Proceedings of the 54th Hawaii international conference on system sciences. University of Hawai'i at Manoa, Hamilton Library

Blaß M, Gimpel H, Graf-Drasch V (2023) When to gather your smartest people in virtual teams – a Mixed-Methods study. Working paper

Bostrom RP, Anson R, Clawson VK (1993) Group facilitation and group support systems. Group Support Syst New Perspect 25:146–148

Bronikowska M, Korcz A, Krzysztoszek J (2019) How years of sport training influence the level of moral competences of physical education and sport students. BioMed Res Intl
Buhrmester M, Kwang T, Gosling SD (2016) Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality data?
Chan J, Dang S, Dow SP (2016) Improving crowd innovation with expert facilitation. In: Gergle D, Morris MR, Bjørn P, Konstan J (eds) Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing. ACM, New York, NY, USA, pp 1223–1235
Chapman KJ, Meuter M, Toy D, Wright L (2006) Can't we pick our own groups? The influence of group selection method on group dynamics and outcomes. J Manag Educ 30:557–569. https://doi.org/10.1177/1052562905284872
Chaves AP, Gerosa MA (2019) How should my chatbot interact? A survey on human-chatbot interaction design. Int J Hum Compu Interact 37:729–758. https://doi.org/10.1080/10447318.2020.1841438
Chen M, Liou Y, Wang CW, Fan YW, Chi YPJ (2007) TeamSpirit: design, implementation, and evaluation of a Web-based group decision support system. Decis Support Syst 43:1186–1202. https://doi.org/10.1016/j.dss.2005.07.008
Clawson VK, Bostrom RP (1996) Research-driven facilitation training for computer-supported environments. Group Decis Negot 5:7–29. https://doi.org/10.1007/BF00554235
Daderko P (2023) The next wave of Copilot in Microsoft Teams—Microsoft Community Hub. https://techcommunity.microsoft.com/t5/microsoft-teams-blog/the-next-wave-of-copilot-in-microsoft-teams/ba-p/3872842. Accessed 28 July 2023
Day EA, Arthur W, Miyashiro B, Edwards BD, Tubre TC, Tubre AH (2004) Criterion-related validity of statistical operationalizations of group general cognitive ability as a function of task type: comparing the mean, maximum, and minimum1. J Appl Soc Psychol 34:1521–1549. https://doi.org/10.1111/j.1559-1816.2004.tb02017.x
Deary IJ (2000) Looking down on human intelligence: from psychometrics to the brain. Oxford psychology series, vol 34. Oxford University Press, Oxford
Dennis AR (1996) Information exchange and use in group decision making: you can lead a group to information, but you can't make it think. MIS Q 20:433. https://doi.org/10.2307/249563
Dennis FV (2008) Media, tasks, and communication processes: a theory of media synchronicity. MIS Q 32:575. https://doi.org/10.2307/25148857
Dennis AR, Wixom BH (2002) Investigating the moderators of the group support systems use with meta-analysis. J Manag Inf Syst 18:235–257. https://doi.org/10.1080/07421222.2002.11045696
Dennis AS, Barlow JB, Dennis AR (2022) The power of introverts: personality and intelligence in virtual teams. J Manag Inf Syst 39:102–129. https://doi.org/10.1080/07421222.2021.2023408
DeSanctis G, Gallupe RB (1987) A foundation for the study of group decision support systems. Manage Sci 33:589–609. https://doi.org/10.1287/mnsc.33.5.589
Devine DJ, Philips JL (2001) Do smarter teams do better. Small Group Res 32:507–532. https://doi.org/10.1177/104649640103200501
Diederich S, Alfred BB, Lutz KM (2020) Designing anthropomorphic enterprise conversational agents. Bus Inf Syst Eng 62:1–17. https://doi.org/10.1007/s12599-020-00639-y
Eden C, Radford J (1990) Tackling strategic problems: the role of group decision support. Sage, London
Engel D, Woolley AW, Jing LX, Chabris CF, Malone TW (2014) Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. PLoS ONE 9:e115212. https://doi.org/10.1371/journal.pone.0115212
Engel D, Woolley AW, Aggarwal I, Chabris CF, Takahashi M, Nemoto K, Kaiser C, Kim YJ, Malone TW (2015) Collective intelligence in computer-mediated collaboration emerges in different contexts and cultures. In: Begole B, Kim J, Inkpen K, Woo W (eds) Proceedings of the 33rd annual ACM conference on human factors in computing systems—CHI '15. ACM Press, New York, pp 3769–3778
French S (2013) Cynefin, statistics and decision analysis. J Oper Res Soc 64:547–561. https://doi.org/10.1057/jors.2012.23
Fuller RM, Dennis AR (2009) Does fit matter? The impact of task-technology fit and appropriation on team performance in repeated tasks. Inf Syst Res 20:2–17
Galletta DF, Zhang Y (2014) Asynchronous virtual teams: can software tools and structuring of social processes enhance performance? Human-computer interaction and management information systems: applications. Adv Manag Inf Syst 135–158
Galluch PS, Grover V, Thatcher JB (2015) Interrupting the workplace: examining stressors in an information technology context. J Assoc Inf Syst 16(1):2

GfeW - German Association for Experimental Economic Research e.V. (2021) Institutional review board-Certificate No. 7za14La8. Available at http://www.gfew.de/ethik/7za14La8
Gimpel H (2015) Interview with Thomas W. Malone on "Collective Intelligence, Climate Change, and the Future of Work." Bus Inf Syst Eng 57(4):275–278
Gimpel H, Graf-Drasch V, Laubacher RJ, Wöhl M (2020) Facilitating like Darwin: supporting cross-fertilisation in crowdsourcing. Decis Support Syst 132:113282. https://doi.org/10.1016/j.dss.2020.113282
Gimpel H, Graf-Seyfried V, Laubacher R, Meindl O (2023) Towards artificial intelligence augmenting facilitation: AI affordances in macro-task crowdsourcing. Group Decis Negot. https://doi.org/10.1007/s10726-022-09801-1
Gimpel H, Graf-Drasch V (2020) Media synchronicity and collective intelligence
Gimpel H, Graf-Drasch V (2023) The multidimensional structure of collective intelligence in virtual groups. Working paper
Gottfredson LS (1997) Mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. Intelligence 24:13–23
Graf-Drasch V, Gimpel H, Barlow JB, Dennis AR (2022) Task structure as a boundary condition for collective intelligence. Pers Psychol 75:739–761. https://doi.org/10.1111/peps.12489
Gu W, Moustafa A, Ito T, Zhang M, Yang C (2021) A case-based reasoning approach for supporting facilitation in online discussions. Group Decis Negot 30:719–742. https://doi.org/10.1007/s10726-021-09731-4
Gupta P, Kim YJ, Glikson E, Woolley AW (2019) Digitally nudging team processes to enhance collective intelligence. In: Collective intelligence 2019
Hauser DJ, Schwarz N (2016) Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav Res Methods 48:400–407. https://doi.org/10.3758/s13428-015-0578-z
Helquist JH, Deokar A, Meservy T, Kruse J (2011) Dynamic collaboration. SIGMIS Database 42:95–115. https://doi.org/10.1145/1989098.1989104
Hendy NT, Biderman MD (2019) Using bifactor model of personality to predict academic performance and dishonesty. Int J Manag Educ 17:294–303
Ito T, Hadfi R, Suzuki S (2022) An agent that facilitates crowd discussion. Group Decis Negot 31:621–647. https://doi.org/10.1007/s10726-021-09765-8
Ito T (2018) Towards agent-based large-scale decision support system: the effect of facilitator. In: Proceedings of the 51st Hawaii international conference on system sciences. University of Hawai'i at Manoa, Hamilton Library
Jonassen DH (1997) Instructional design models for well-structured and ill-structured problem-solving learning outcomes. Educ Tech Res Dev 45:65–94. https://doi.org/10.1007/bf02299613
Khalifa M, Kwok R-W, Davison R (2002) The effects of process and content facilitation restrictiveness on GSS-mediated collaborative learning. Group Decis Negot 11:345–361. https://doi.org/10.1023/A:1020449317854
Kilgour DM, Eden C (eds) (2010) Handbook of group decision and negotiation, vol 4. Springer, Dordrecht
Kim YJ, Engel D, Woolley AW, Lin JY-T, McArthur N, Malone TW (2017) What makes a strong team? In: Lee CP, Poltrock S, Barkhuus L, Borges M, Kellogg W (eds) Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, pp 2316–2329
Kim S, Eun J, Oh C, Suh B, Lee J (2020) Bot in the Bunch: facilitating group chat discussion by improving efficiency and participation with a Chatbot. In: Conference on human factors in computing systems—proceedings. Association for Computing Machinery, pp 1–13
Kitchner KS (1983) Cognition, metacognition, and epistemic cognition. Hum Dev 26:222–232. https://doi.org/10.1159/000272885
Kittur A, Lee B, Kraut RE (2009) Coordination in collective intelligence. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 1495–1504. https://doi.org/10.1145/1518701.1518928
Kolfschoten GL, Briggs RO, de Vreede G-J, Jacobs PH, Appelman JH (2006) A conceptual foundation of the thinkLet concept for Collaboration Engineering. Int J Hum Comput Stud 64:611–621. https://doi.org/10.1016/j.ijhcs.2006.02.002
Landauer TK (1987) Psychology as a mother of invention. ACM SIGCHI Bull 18:333–335. https://doi.org/10.1145/1165387.275653

Lim J, Guo X (2008) A study of group support systems and the intergroup setting. Decis Support Syst 45:452–460. https://doi.org/10.1016/j.dss.2007.06.007

Limayem M, DeSanctis G (1993) Automating decision guidance in a group decision environment

Limayem M, Lee-Partridge JE, Dickson GW, DeSanctis G (1993) Enhancing GDSS effectiveness: automated versus human facilitation. In: Proceedings of the twenty-sixth Hawaii international conference on system sciences, pp 95–101. https://doi.org/10.1109/hicss.1993.284171

Lind G (2000) Moral regression in medical students and their learning environment. Revista Brasileira De Educacao Médica 24:24–33

Lind G, Hartmann H (1985) Testing for moral judgment competence. In: Wakenhut R (ed) Moral judgments and social education. Routledge, New York, pp 79–105

Malone TW, Crowston K (1990) What is coordination theory and how can it help design cooperative work systems? In: Proceedings of the 1990 ACM conference on computer-supported cooperative work, pp 357–370

Malone TW, Crowston K (1994) The interdisciplinary study of coordination. ACM Comput Surv 26:87–119. https://doi.org/10.1145/174666.174668

McGrath JE (1984) Groups, interaction and performance. Prentice-Hall, Englewood Cliffs

Meslec N, Aggarwal I, Curseu PL (2016) The insensitive ruins it all: compositional and compilational influences of social sensitivity on collective intelligence in groups. Front Psychol 7:676. https://doi.org/10.3389/fpsyg.2016.00676

Moradi Dakhel A, Majdinasab V, Nikanjam A, Khomh F, Desmarais MC, Jiang ZM (2023) GitHub Copilot AI pair programmer: asset or liability? J Syst Softw 203:111734. https://doi.org/10.1016/j.jss.2023.111734

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs

Newman SA, Dynamics RF (2021) Five steps to leading your team in the virtual COVID-19 workplace. Organ Dyn 50:100802

Nunamaker JF, Dennis AR, Valacich JS, Vogel D, George JF (1991) Electronic meeting systems. Commun ACM 34:40–61. https://doi.org/10.1145/105783.105793

Nunamaker JF, Briggs RO, Mittleman DD, Vogel DR, Pierre BA (1996) Lessons from a dozen years of group support systems research: a discussion of lab and field findings. J Manag Inf Syst 13:163–207. https://doi.org/10.1080/07421222.1996.11518138

O'Brien GE, Owens AG (1969) Effects of organizational structure on correlations between member abilities and group productivity. J Appl Psychol 53:525–530. https://doi.org/10.1037/h0028659

Ostrowski B, Woolley AW, Haan K (2019) Group brainstorming: the effects of collective intelligence, individual ability, and task structure. Acad Manag. https://doi.org/10.5465/AMBPP.2019.13905abstract

Pervan GP (1998) A review of research in Group Support Systems: leaders, approaches and directions. Decis Support Syst 23:149–159. https://doi.org/10.1016/S0167-9236(98)00041-4

Porter CM, Woo SE, Allen DG, Keith MG (2019) How do instrumental and expressive network positions relate to turnover? A meta-analytic investigation. J Appl Psychol 104:511–536. https://doi.org/10.1037/apl0000351

Przybilla L, Baar L, Wiesche M, Krcmar H (2019) Machines as teammates in creative teams. In: Joseph D, van Slyke C, Allen JP, Quesenberry J, Wiesche M (eds) Proceedings of the 2019 on computers and people research conference. ACM, New York, pp 94–102

Rakhmanov O, Dane S (2021) Article in communications in computer and information science. In: Information and communication technology and applications: third international conference, ICTA 2020, Minna, Nigeria, November 24–27, 2020, Revised Selected Papers 3:54–65. https://doi.org/10.1007/978-3-030-69143-1_5

Rico R, Sánchez-Manzanares M, Gil F, Gibson C (2008) Team implicit coordination processes: a team knowledge-based approach. Acad Manag Rev 33:163–184. https://doi.org/10.5465/amr.2008.27751276

Riedl C, Kim YJ, Gupta P, Malone TW, Woolley AW (2021) Quantifying collective intelligence in human groups. Proc Natl Acad Sci USA 118:e2005737118. https://doi.org/10.1073/pnas.2005737118

Rissler R (2017) The impact of IT-mediated interruptions on Flow and task performance. In: Doctoral consortium proceedings of the 25th European conference on information systems (ECIS), Guimarães, Portugal, pp 5–10

Rowe L (2019) Exploring collective intelligence in human groups. Dissertation, The University of Melbourne

Schraw G, Dunkle ME, Bendixen LD (1995) Cognitive processes in well-defined and ill-defined problem solving. Appl Cogn Psychol 9:523–538. https://doi.org/10.1002/acp.2350090605
Shewach OR, Sackett PR, Quint S (2019) Stereotype threat effects in settings with features likely versus unlikely in operational test settings: a meta-analysis. J Appl Psychol 104:1514–1534. https://doi.org/10.1037/apl0000420
Simon HA (1973) The structure of ill structured problems. Artif Intell 4:181–201
Snowden DJ, Boone ME (2007) A leader's framework for decision making. Harv Bus Rev 85:68
Spearman C (1904) General intelligence' objectively determined and measured. Am J Psychol 15:201–293
Stallbaumer C (2023) Introducing Microsoft 365 Copilot—a whole new way to work. Microsoft 365 Blog
Stasser G, Titus W (2003) Hidden profiles: a brief history. Psychol Inq 14:304–313. https://doi.org/10.1080/1047840X.2003.9682897
Steiner ID (1972) Group process and productivity. Academic Press, New York
Stern A, Schultze T, Schulz-Hardt S (2017) How much group is necessary? Group-to-individual transfer in estimation tasks. Collabra Psychol. https://doi.org/10.1525/collabra.95
Tavanapour N, Bittner EA (2018) Automated facilitation for idea platforms: design and evaluation of a chatbot prototype
Tews MJ, Michel JW, Noe RA (2011) Beyond objectivity: the performance impact of the perceived ability to learn and solve problems. J Vocat Behav 79:484–495
Thissen MR, Page JM, Bharathi MC, Austin TL (2007) Communication tools for distributed software development teams. In: SIGMIS-CPR 2007—proceedings of the 2007 ACM SIGMIS CPR conference: the global information technology workforce, pp 28–35
Valacich JS, Jung JH, Looney CA (2006) The effects of individual cognitive ability and idea stimulation on idea-generation performance. Group Dyn Theory Res Pract 10:1–15. https://doi.org/10.1037/1089-2699.10.1.1
Wilson J, Rosenberg D (1988) Rapid prototyping for user interface design. In: Handbook of human-computer interaction, pp 859–875
Winkler R, Roos J (2019) Bringing AI into the classroom: designing smart personal assistants as learning tutors. In: Proceedings of the 40th international conference on information systems
Wonderlic EF (1992) Wonderlic personnel test user's manual
Wong Z, Aiken M (2003) Automated facilitation of electronic meetings. Inf Manag 41:125–134. https://doi.org/10.1016/S0378-7206(03)00042-9
Woolley AW, Chabris CF, Pentland A, Hashmi N, Malone TW (2010) Evidence for a collective intelligence factor in the performance of human groups. Science 330:686–688. https://doi.org/10.1126/science.1193147
Woolley AW, Riedl C, Kim YJ, Malone TW (2017) More evidence for a general collective intelligence factor in human groups: a meta-analysis. In: Collective intelligence 2017
Yuan A, Coenen A, Reif E, Ippolito D (2022) Wordcraft: story writing with large language models. In: 27th international conference on intelligent user interfaces. Association for Computing Machinery, New York, pp 841–852
Zigurs I, Buckland BK (1998) A theory of task/technology fit and group support systems effectiveness. MIS Q 22:313. https://doi.org/10.2307/249668

Authors and Affiliations

Henner Gimpel[1,2,3] · **Stefanie Lahmer**[1,2] · **Moritz Wöhl**[1,2] · **Valerie Graf-Drasch**[1,2,3]

✉ Stefanie Lahmer
stefanie.lahmer@fit.fraunhofer.de

Henner Gimpel
henner.gimpel@uni-hohenheim.de

Moritz Wöhl
moritz.woehl@fit.fraunhofer.de

Valerie Graf-Drasch
valerie.graf-drasch@uni-hohenheim.de

1 Branch Business and Information Systems Engineering, University of Hohenheim, FIM Research Center, Fraunhofer Institute for Applied Information Technology FIT, Schloß Hohenheim 1, 70599 Stuttgart, Germany

2 Branch Business and Information Systems Engineering, Fraunhofer Institute for Applied Information Technology FIT, Alter Postweg 101, 86159 Augsburg, Germany

3 University of Hohenheim, Schloß Hohenheim 1, 70599 Stuttgart, Germany